Cha, Junseo; Choi, Seong Hee; Choi, Chul-Hee

¿Se puede mejorar una voz de canto natural a través del procesamiento digital? Implicaciones del entrenamiento de la voz y la vocología en cantantes

Revista de investigación e innovación en ciencias de la salud, vol. 3, núm. 2, 2021, Julio-Diciembre, pp. 72-86

Fundación Universitaria María Cano

DOI: https://doi.org/10.46634/riics.119

REFLECTION ARTICLE

OPEN ACCESS

REVISTA RIICS Investigación e innovación en ciencias de la salud

Can a natural singing voice be enhanced through digital processing? Implications of voice training and vocology in singers

¿Se puede mejorar una voz de canto natural a través del procesamiento digital? Implicaciones del entrenamiento de la voz y la vocología en cantantes

Junseo Cha[1], **Seong Hee Choi**[1, 2, 3], **Chul-Hee Choi**[1, 2, 3]

[1] *Department of Audiology and Speech-Language Pathology; The Graduate School; Daegu Catholic University; Gyeongsan; South Korea.*

[2] *Research Institute of Biomimetic Sensory Control; Daegu Catholic University; Gyeongsan; South Korea.*

[3] *Catholic Hearing Voice Speech Center; Daegu Catholic University; Gyeongsan; South Korea.*

Correspondence
Chul-Hee Choi. Department of Audiology & Speech-Language Pathology and Research Institute of Biomimetic Sensory Control, and Catholic Hearing Voice Speech Center, Daegu Catholic University, 13-13 Hayang-ro, Hayang-up, Gyeongsan-si, Gyeongsanbuk-do, 712-702, Korea. Email: cchoi@cu.ac.kr

How to cite
Cha, Junseo; Choi, Seong Hee; Choi, Chul-Hee. (2021). Can a natural singing voice be enhanced through digital processing? Implications of voice training and vocology in singers. Revista de Investigación e Innovación en Ciencias de la Salud. 3(2), 72-86. https://doi.org/10.46634/riics.119

Received: 29/10/2021
Revised: 14/11/2021
Accepted: 18/11/2021

Invited editor
Carlos Manzano Aquiahuatl, MD, MSc.

Editor en jefe
Jorge Mauricio Cuartas Arias, Ph.D.

Coeditor
Fraidy-Alonso Alzate-Pamplona, MSc.

Abstract

Introduction. The traditional way of facilitating a good singing voice has been achieved through rigorous voice training. In the modern days, however, there are some aspects of the singing voice that can be enhanced through digital processing. Although in the past, the frequency or intensity manipulations had to be achieved through the various singing techniques of the singer, technology today allows the singing voice to be enhanced from the instruments within recording studios. In essence, the traditional voice pedagogy and the evolution of digital audio processing both strive to achieve a better quality of the singing voice, but with different methods. Nevertheless, the major aspects of how the singing voice can be manipulated are not communicated among the professionals in each field.

Objective. This paper offers insights as to how the quality of the singing voice can be changed physiologically through the traditional ways of voice training, and also digitally through various instruments that are now available in recording studios.

Reflection. The ways in which singers train their voice must be mediated with the audio technology that is available today. Although there are aspects in which the digital technology can aid the singer's voice, there remain areas in which the singers must train their singing system in a physiological level to produce a better singing voice.

Keywords

Voice pedagogy; voice training; singing; vocology; voice science; contemporary commercial music; audio engineering; frequency response; equalizer; dynamic range compressor.

Conflicts of Interest
The authors have declared that no competing interests exist.

Data Availability Statement
All relevant data is in the article and in the appendices. For more detailed information, write to the Corresponding Author.

Funding
None. This research did not receive any specific grants from funding agencies in the public, commercial, or non-profit sectors.

Disclaimer
The content of this article is the sole responsibility of the authors and does not represent an official opinion of their institutions or the Revista de Investigación e Innovación en Ciencias de la Salud.

Author Contributions
Junseo Cha: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft, writing – review & editing.
Seong Hee Choi: conceptualization, investigation, methodology, resources, software, supervision, writing – original draft, writing – review & editing.
Chul-Hee Choi: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft, writing – review & editing.

Resumen

Introducción. La forma tradicional de facilitar una buena voz para cantar se ha logrado mediante un riguroso entrenamiento de la voz. Sin embargo, en la actualidad, existen aspectos de la voz cantada que pueden mejorarse mediante el procesamiento digital. Aunque en el pasado las manipulaciones de frecuencia o intensidad tenían que lograrse a través de las diversas técnicas de canto del cantante, la tecnología actual permite ahora mejorar la voz del canto desde los instrumentos dentro de los estudios de grabación. En esencia, la pedagogía de la voz tradicional y la evolución del procesamiento de audio digital se esfuerzan por lograr una mejor calidad de la voz cantada, pero con métodos diferentes. No obstante, los principales aspectos de cómo se puede manipular la voz cantada no se comunican entre los profesionales de cada campo respectivo.

Objetivo. Este artículo ofrece información sobre cómo la calidad de la voz cantada se puede cambiar fisiológicamente a través de las formas tradicionales del entrenamiento de la voz, y también digitalmente a través de varios instrumentos que ahora están disponibles en los estudios de grabación.

Reflexión. Las formas en que los cantantes entrenan su voz deben estar mediadas por la tecnología de audio que está disponible en la actualidad. Aunque hay aspectos en los que la tecnología digital puede ayudar a la voz del cantante, quedan áreas en las que los cantantes deben entrenar su sistema de canto a nivel fisiológico para producir una mejor voz al cantar.

Palabras clave

Pedagogía de la voz; entrenamiento de la voz; canto; vocología; ciencia de la voz; música comercial contemporánea; ingeniería de audio; respuesta frecuente; ecualizador; compresor de rango dinámico.

Introduction

According to Nielsen and Billboard, the most streamed music genre in United States in 2020 were consecutively R&B/Hip-Hop, Rock, and Pop genres, amassing a total of 59.8%, while Classical genre consumption were recorded as 0.8% [1]. While classical music started losing audience from the last half of the 20th century [2], contemporary commercial music (CCM) has gained more popularity among avid music listeners.

Moreover, the format of music in which listeners purchase music has changed dramatically over the last half century. According to the Recording Industry Association of America (RIAA), the format of music has significantly changed from analog into the digital realm from 1973 to 2020. While in the 70's the predominant formats of music consumers were LP/EP and cassette tapes, most music revenues in 2020 are through paid online subscription in music streaming platforms [3]. This has made music more accessible for listeners without the need to go to a concert hall or to purchase an expensive audio playback device. In addition, consumers with personal computers and smartphones with a music player device can easily purchase and listen to the music.

During the platform change of music consumption, the production of music was in the midst of being revolutionized as well. From producing music in a recording

studio with an 8-track tape machine, the evolution of digital technology has made possible to make music with computer software called Digital Audio Workstation (DAW) and Virtual Studio Technology (VST) [4]. A DAW called Cubase (manufactured by Steinberg) was introduced in 1989 initially as a sequencer program to write digital information into a computer in a certain sequence. After the user selects a certain instrument, the sequencer would emulate such a sound within the program [5]. Meanwhile, another DAW Pro Tools (manufactured by Digidesign) was introduced in 1991 and was being used to record instruments [6], potentially replacing the expensive and voluminous 48-channel audio consoles and tape machines with a single computer in recording studios.

Although dramatic changes have developed in the recording industry, the traditional voice training methods for professional singers by voice teachers and coaches to produce a better singing voice have stayed relatively stagnant. Therefore, instead of depending only on the experiences of voice teachers, voice training based on scientific rationales should be developed and established for singers.

In addition, audio engineers try to develop different techniques to control or manipulate the frequency response and intensity through the modern recording systems, including microphones, preamplifiers, analog to digital converters, digital audio workstations, virtual studio devices, digital to analog converters, and loudspeakers. They have the same goal as the professional trainers of singing voice but use different techniques. The engineers achieve the goals through manipulating the frequency responses of different instruments such as microphones, preamplifiers, and equalizers, using vocal equalization techniques and multiple speaker systems, reducing some distortions, and using compressions.

The voice training field for professional singers and the voice processing field of the audio engineers share the same goals to improve the quality of the singing voice, but so far there is not much communication between those fields. This paper provides meaningful insights on how these professionals communicate with each other and encourage professionals in either field to discuss the major common topics related to the voice training/processing methods for singers and engineers.

Voice Training Methods for Singers

Voice Training Techniques to Improve the Voice Timbre

Singing is defined as the act of producing musical sounds with the human's musical instrument: the voice [7]. Generally, professional singers in opera or vocalists in jazz and popular music have gotten voice training provided by voice teachers and coaches throughout their careers to produce a good singing voice. A good singing voice is related to a lot of the acoustical characteristics of voice, including voice types, vocal range, vocal weight, vocal tessitura, vocal timbre, and voice transition points, which can be obtained from properly using or manipulating the organs related to voice production.

The organs related to voice production consists of the lungs functioning as an air supplier, the larynx acting as a vibrator, the vocal tract functioning as resonators, and the tongue, palate, teeth, and lips acting as articulators. Each component of the vocal organs functions independently or coordinates with one another to produce a singing voice. Although it is important that singers have a natural gift to produce a good sound, they should still learn to make their voice mechanisms coordinate in certain ways unique to their vocal organs.

In terms of vocal training and habilitation for singers, supraglottic activity is known as one of the important factors contributing to sound quality [8,9]. The vocal tract above the glottis can be manipulated in shape by the singers to change the resonant frequencies, thus enhancing the intensity in a certain range of frequencies of the singing voice. The resonant peaks shown in the spectral output are known as formants, and singers control their vocal tract shape to enhance different frequencies of the voice, thus making the timbre of the voice darker or brighter. A particular technique in formant repositioning is widely used among singers to make their singing voice brighter and easily heard among the other instruments that constitute a song. This goal can be achieved by the use of the "singer's formant", which enhances the voice at 3 kHz to be better perceived when singing with the orchestra [10]. The singer's formant indicates a formant cluster of the third, fourth, and fifth formants around 3 kHz of the voice. This may cause the difference in the acoustic properties between the spectrum of the singing voice and the orchestra. The orchestra produces a lot of energy around 500 Hz, but it falls steadily at high frequencies and produces relatively little energy at 3,000 Hz. On the other hand, a well-trained operatic voice can be heard when sung with the orchestra, because it produces a concentration of energy at 3,000 Hz. According to a study by Sundberg (1974), the cause of a strong vocal resonance is possible by changing the area and the shape of a small resonator inside the vocal tract just above the vocal folds called the 'epilarynx', which allows it to produce the singer's formant. For optimal resonance, this resonator should be 1/6 long in length of the total resonator and also should have 1/6 of the area of the total resonator. The narrowed epilaryngeal tube is a virtual source of the "singer formula" that increases spectral energy of about 2.6-3.0 kHz. It is also associated with an increased perception of "ring". Another explanation of the singer's formant is based on the overlap of voice and hearing ranges in singing [11]. Acoustically, the human's ears display the resonant frequency of 2.6-3.0 kHz with maximal gain [12]. The trained singers boost the resonant frequency of the ears so that the audience with normal hearing can perceive their singing voices among the spectrum of the orchestra [11]. Therefore, this indicates that the trained singer's voice is not only enhanced acoustically by the singer's formant, but also perceptually because the singer's formant represents the most sensitive frequency ranges in the auditory system. On the other hand, this also indicates that all vocalists appear to be perceptually weak at the lower and higher ends of the orchestra spectrum, which have less harmonic energy and less sensitive frequency ranges. Therefore, singers have somehow learned to maximize the sound perception of listeners with normal hearing by changing the level of spectral peaks in their voices in orchestral environments.

The epilarynx constriction facilitates vocal fold vibration, increases inertance, and improves impedance matching with glottis to reduce the phonation threshold pressure [13,14]. Moreover, increased inertial response also leads to the increase in the maximum flow declination rate (MFDR) and harmonic amplitude [14] and causes perceivable changes in the voice quality [15]. Thus, this provides the underlying mechanism of a variety of common voice training or therapeutic techniques such as resonant voice therapy [16]. Titze described resonant voice and defined it as any voice production that is both easy to produce and creates vibratory sensations in the face [17]. Acoustically, it is an indication of effective conversion of aerodynamic energy to acoustic energy, rather than the sound resonating in the sinuses or the nasal airways and perceptually as barely adducted (not pressed) or barely abducted (not breathy) [18]. In addition, in terms of vocal pedagogy, video laryngoscopic findings have demonstrated aryepiglottic constriction for twang, belting, and opera singing voice qualities [19].

Recently, there has been studies that show that the extension of the vocal tract with a tube, a straw, a megaphone, or other device can be beneficial in "training" the sound source for less-effortful production with greater efficiency [20]. Singing with a narrow tube at the end of the lips is considered to be one of the exercises conducted with a semi-occluded vocal tract (SOVTEs). In a computer modeling, /u/vowel was simulated with different cross-sectional areas of the epilarynx tube and artificially lengthened with the "resonance tube" [21]. They found that the effect was strong when the epilarynx tube was narrowed along with the expansion of the vocal tract with the "resonance tube". This phenomenon suggests that singers can optimize their voice by modifying their vocal tract configuration, including a narrow epilarynx. Therefore, phonation into a tube suggest that vocal trainee may help find the optimal glottal and epilaryngeal setting for the greatest vocal economy. In practice, voiced obstruents and phonation into tubes with the end of the tubes in free air or immersed in water are widely used as vocal exercises. From this point of view, SOVTEs such as humming, lip or tongue trills, and sustained voiced fricatives, and tube or straw phonation have been beneficial for the improvement of the functionally singing [14,22,23].

Moreover, trained singers can change the timbre of their voice through first and second formant repositioning. This can be achieved by changing the resonances of the vocal tract. All formant frequencies decrease uniformly as the length of the vocal tract increases and decreases uniformly with lip rounding and increases with lip spreading [24]. In practice, singers can control their articulators so that they can shift the shape of their resonators and thus produce different timbres that fit the song's needs. For example, operatic singers often use vowel modifications to maintain a uniform timbre of sound throughout their frequency ranges as well as to loudly project their voices, and still remain understood by the audience [25]. When the singers modify the vowels to approximate an inverted megaphone mouth shape, the back of the mouth and pharynx were wide with only a moderate lip opening, which is as often taught in classical singing in the G4–D5 pitch range. Although there is little difference in the way males and females approach vowel modification in the G4–D5 range, the modifications are highly dependent on different singing styles. The fundamental frequency (f_o) appears to be reinforced by F_1 for classical singing, whereas the second harmonic ($2f_o$) appears to be reinforced by F_1 for belting [26]. Since vowel modification affects vowel intelligibility along with voice timbre, it is a part to consider in voice pedagogy.

Voice Training Techniques to Increase the Voice Intensity

Producing a better singing voice can also be achieved by intensifying the singing voice without applying excessive damage to the tissue of the vocal folds. The intensity of the singing voice is related to the pressure of the singing voice. Acoustically, intensity is proportional to the square of pressure. The relative intensity level is obtained by the dB IL formula ($dB\ IL = 10 \log \frac{Ix}{Ir}$), displaying the ratio of the intensity level (Ix) of the singing voice in terms of the reference intensity level (Ir), while the pressure level is obtained by the dB SPL formula ($dB\ SPL = 20 \log \frac{Px}{Pr}$), displaying the ratio of the sound pressure level (Px) of the singing voice in terms of the reference sound pressure level (Pr). Based on the dB IL and dB SPL formula, if the intensity level is double than the reference intensity level, it results in an increase of 3 dB IL, while if the sound pressure level is double than the reference sound pressure level, it leads to an increase in 6 dB SPL. Singers need to double the acoustic power to increase the 3 dB IL or 6 dB SPL in the output of their singing voice. This indicates that it is important for singers to learn how to increase the power of their resonators to produce a higher intensity of the singing voice.

Several methods to produce a higher intensity of the singing voice have been discussed in the realm of voice training for singers. The first way is the use of the subglottal lung pressure. When comparatively analyzing the subglottal waveforms of the five professional tenor and 25 normal subjects, the professional singers showed the increased fundamental frequency at a rate of 8-9 dB/octave, due to the higher subglottal lung pressure and produced 10-12 dB greater intensity than the untrained singers [27]. This indicates that the voice intensity changes as a function of the fundamental frequency and the subglottal lung pressure [27,28].

The second way is through the adduction of the vocal folds and the airway shape. The relative contribution of the vocal fold adduction and airway shape on the loudness control of singers is less known. Although the loudness changes as the glottal width narrows to a certain point, the airway shapes of the vocal tract seem to change the loudness relatively more. When the airway forms are changed from a uniform tube to a belt or a nominal shape, the loudness increased with a slight decrease in SPL; when the airway forms are changed from a uniform tube to an operatic ring shape, the loudness increased with a small increase in SPL [29].

Recently, there is also a study that shows vocalizations with a semi-occluded airway have an overall widening effect on the airway and can lead to variations in sound intensity, resulting in optimization in sound production [20]. When quantifying the steady airflow resistances in a semi-occluded airway as well as acoustic impedances in vocalization from the lungs to the lips, vocalizations with a semi-occluded airway can regulate the laryngeal airflow resistance and maximize the aerodynamic power transfer by matching the impedance of the source to the impedance of the vocal tract [30].

In addition, the vocal tract shapes can affect the frequency responses of the singing voice. When comparing the different vocal tract shapes such as the operatic singing, using vowels that are modified toward an inverted megaphone mouth shape when singing a high-pitch sound and the jazz or theater belting techniques using vowels that are consistently modified toward the megaphone (trumpet-like) mouth shape, the vocal tract shapes provide reinforcement to multiple harmonics in the form of inertive supraglottal reactance and compliant subglottal reactance [31]. Especially, the operatic singing allows all the harmonics, except the fundamental, to be "lifted" over the first formant while the belting technique maintains both the fundamental and the second harmonics below the first formant [31]. For clinical and voice training applications, the primary focus is on the two airway conditions, an oral semi-occlusion and a semi-occlusion above the vocal folds [30].

In voice pedagogy, singers try to master epilaryngeal airway control and this efficiency can be maintained when the mouth is opened after practice. Meanwhile, a semi-occlusion at the lips helps to set up this condition. Thus, tube or straw phonation can provide acoustic and aerodynamic benefits for untrained singers. In addition, voice production with a semi-occluded vocal tract increases voice efficiency, which allows singers to sing in a loud voice with less effort. In particular, using relatively an excessive amount of lung pressure during semi-occlusions may present little voice problems because greater lung pressure in this condition does not always produce more vocal folds collision [30]. With semi-occluded vocalization with tube or straw, increased back pressure, which reflects at the lips and back to the vocal folds, causes less vocal fold collision.

Voice Processing Methods for Singers in Recording Studios

Signal Processing Procedures in Recording Studios

When recording the voice in recording studios, the modern recording system consists of a lot of electrical or acoustic instruments such as microphones, preamplifiers, analog to digital converters, digital audio workstations, virtual studio devices, digital to analog converters, and loudspeakers. These instruments are connected as a sequence of the input and the output as shown in Figure 1. The singing voice enters the microphone as an input and the output of microphone goes to the input of the preamplifier connecting to the analog to digital converter, digital audio workstation, virtual studio devices, digital to analog converter, and loudspeakers, respectively. The microphone converts an acoustic signal to an electrical signal that is then intensified by the preamplifier. When the intensified electrical signal passes through an "Analog-to-Digital Converter", it is converted into a digital signal with a sampling rate and bit depth set through sampling, quantization, and encoding. The converted digital information is shown as an audio waveform within a Digital Audio Workstation (DAW).

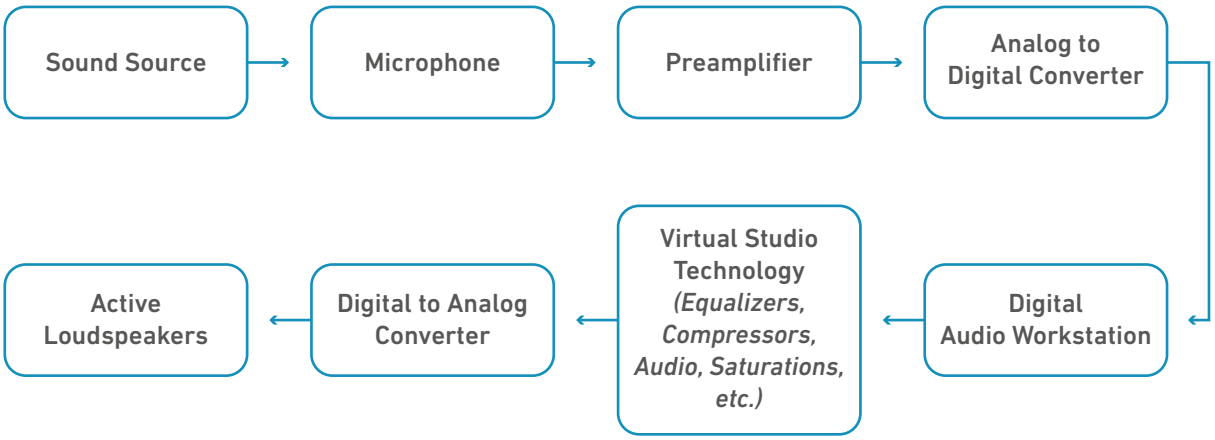

Figure 1. Signal Processing Procedures of Voice Production in Recording Studios.

The stages from "Digital Audio Workstation" to "Active Loudspeakers" are called audio mixing, the process of manipulating each audio track and combining them into a complete stereo track. At these stages, engineers can freely choose any audio tracks within the DAW and adjust its sound through Virtual Studio Technology (VST) processing. During the processing, the engineers use VST effects to change the recorded sounds within the DAW. Equalizers, dynamic audio compressors, and pitch correction tools are some examples of VSTs that are applied to the vocal tracks frequently nowadays, as they seem to play a considerable factor in enhancing the quality of the singing voice. The number of VST effects that can be used in the DAW can be limitless. The engineers can monitor the audio manipulations made by the VST effects through the pair of loudspeakers in the studio. During the monitoring process, the digital audio information is passed through a "Digital-to-Analog Converter" and is transferred into an electrical signal, which is then amplified and converted into an acoustic signal by the active loudspeakers. When the engineers are satisfied with the sound that comes out of the pair of loudspeakers, the multiple audio tracks within the DAW are exported as a single stereo audio file. The engineer can then choose to process more VST effects to the combined stereo track to manipulate the audio further.

After the mixing process is finished, the stereo audio track is manipulated during the mastering process. The mastering engineer applies VST effects such as limiters, multiband compressors, dynamic equalizers, or mid-side audio equalizers to further balance the audio's frequency regions and achieve more loudness. The limiter is usually applied so that the loudest part of the song reaches 0.0 dBFS or -0.1 dBFS, and the overall loudness is dependent on the genre of the song. After the mastering process, the final processed audio is distributed to streaming platforms.

Techniques for Manipulating Frequencies of the Singing Voice in Recording Studios

The audio engineers in recording studios put a lot of effort in producing a better quality of the singing voice through manipulation of the frequency response of the acoustical or electrical instruments, as shown in Figure 1. First, the engineers take into account the frequency responses of microphones. They can be manipulated by controlling the maximal output and the gain of the microphones. It is important to recognize that different types of microphones have different frequency responses to a singing voice. The most frequently used microphones in recording studios are condenser and dynamic microphones. Condenser microphones consist of two plates (a diaphragm and a back plate) separated by a very small distance and form a capacitor, while the dynamic microphone composes of a diaphragm attached to a small coil of wire fitted around the core of a magnet [32]. Condenser microphones can have a very flat frequency response extending from 20 to 20,000 Hz, which is the human ear's frequency range, while dynamic microphones have a resonance peak at about 3 to 6 kHz that can be absent in the condenser microphone. When comparing the condenser and dynamic microphones, the condenser microphones can more accurately capture sound waves across the whole spectrum due to its sensitivity and structural design than dynamic microphones, although condenser microphones are usually rather expensive. In addition, condenser microphones can extend the recording frequencies up to the range of 100 kHz. The actual frequency response of the condenser microphone depends on the construction of the microphone and the physical properties of the diaphragm such as diameter, size, shape, weight, or tensile strength [33].

The audio engineers in recording studios may favor the condenser microphone for recording the singing voice because of the relatively broad and flat frequency response of the microphone. The condenser microphones have a greater advantage when very sensitive measures must be recorded and any deviations in the measured signal caused by the signal's changes must be easily detected. However, although the relatively flat frequency response is technologically viable, it is always not prioritized among engineers to record the singing voice. The engineers can prefer a different sound quality so that the high frequencies are increased. One of the methods to boost the high frequencies is to attach a small amplifier known as a cathode follower to the microphone. This acts as a preamplifier and boosts the output signal in the high frequencies. Another method is to use a large diaphragm microphone. This large diaphragm can also boost acoustic energy at the high frequency regions due to the tensile strength of the large diaphragm. The reason for the high frequency boost during voice recording can be attributed to the resonant frequency of 2.6 kHz of the human ear. If the output of the condenser microphone is closer to the resonant frequency of the human ear, the singing voice can be easily recognized by listeners with normal hearing.

The frequency response from the input microphone cannot be regulated by the engineers as meticulously. After the singing voice has passed through the input of the microphone and

the preamplifier along with the other instruments of the song, the recorded sounds will not be balanced in timbre and volume. While the sounds of some instruments can be heard distinctively, the sounds of other instruments sharing the same frequency domain can be less perceived. At this stage, there is a psychoacoustic process of raising the hearing threshold for one sound (signal) by presentation of another sound (masker), called a making effect [33]. When there are two acoustic signals sharing the same frequency domain and different intensity, the louder sounds (signals) dominate the frequency region and are more perceived while the weaker signals (masker) become "masked" and difficult to perceive. In addition, when the level of a masker tone is above 60 dB at low frequencies, the masking effect of a sound is much greater above the masking frequency than below the masking frequency. The masking effect spreads rapidly upward to higher frequencies as the intensity level of the masker is increased, called upward spread of masking [34]. Due to the upward spread of masking, the lower frequency with high intensity can be heard among the masking noises.

Depending on the genre, a song may be comprised of a minimum of 2 audio tracks all the way up to more than 400 audio tracks, making them susceptible to auditory masking problems. As the track count increases, it creates a chaos of unintelligible glut of audio information that must be balanced and adjusted properly. This balancing process is called mixing. Audio engineers reconcile this issue partly with the use of equalizers. Equalizers have multiple filters that govern the gain of their selected frequency ranges. While the high-pass and the low-pass filter are responsible for determining the cut-off frequencies in each end of the frequency spectrum, the band-pass filters permit the engineers to amplify or attenuate the singing voice in only any chosen frequency ranges. Through equalizations, the resulting output sound of each instrument may become clearer and be more defined in a mixture of other audio tracks [35].

Audio engineers usually seek to emphasize the singing voices to be heard as clear and distinct as possible, while avoiding the vocal track from being masked by other melodic instruments sharing the same frequencies such as pianos, guitars, and string instruments [35]. Equalizers allow the engineers to attenuate the selected frequencies of instruments that collide with the intelligibility of the singing voices. If the original recorded voice itself is less defined, the engineers may boost the higher region of the voice's frequency or attenuate the lower areas to attain more clarity and definition to the voice [36]. While the physiological changes in the vocal organs can result in a different voice timbre through formants repositioning, the use of equalizers now allow the voice to be changed in a more controlled and precise way.

Audio engineers have used some vocal equalization techniques in adjusting the singing voice to a brighter timbre than the natural state of the human voice [36]. The engineers try to boost or emphasize the frequency region of 2~5 kHz, the most sensitive frequency ranges in the human ears, so that the singing voices are more audible compared to the other instruments [36]. This is due to the equal loudness contour displaying the variation in the auditory sensitivity across frequency [34]. The equal loudness contour represents the physical loudness magnitude in which the subjective loudness of other frequencies is equally perceived to the level in dB of the standard 1 kHz tone. Phon is used as the unit of equal loudness contour. The 40 phons indicates the same loudness perceived as 40 dB loudness as a function of frequency. For example, the 40 dB SPL at 1 kHz and approximately 55 dB SPL at 100 Hz are perceived as the same loudness (40 phons). In the equal loudness contour, the differences between the physical sound pressure level in dB and the loudness level in phons are very large at low levels and less at high levels. In addition, the loudness level in phons is least between

2-5 kHz, indicating the most sensitive frequency ranges to the human ears. Therefore, audio engineers focus on boosting at the frequency ranges to make the singing voice more audible and brighter in the mixture of other sounds from other instruments.

While equalizers play a role in manipulating certain frequency regions of the singing voice, pitch correction tools can manipulate the fundamental frequency of the singing voice itself. The fundamental frequency of the voice is known to be physiologically attributed to the interaction between the activation of the thyroarytenoid muscle and the cricothyroid muscle, along with the lung pressure [24]. According to the muscle activation plot by Titze, the sole increase in either the thyroarytenoid muscle or the cricothyroid muscle activity increases the fundamental pitch in a non-linear fashion. Moreover, as the lung pressure increases, it elongates the vocal folds further during its vibratory cycle, resulting in a higher stress of the vocal folds and therefore creates higher pitch. Skilled singers control their fundamental frequency by mediating the function of the thyroarytenoid, cricothyroid, and the lung pressure to produce a controlled pitch during singing [24,28]. However, in practice, it is common for singers to deviate slightly from the targeted pitch during a performance. While some may be imperceptible to the listeners, there may be instances where out-of-tune singing can be clearly heard. Pitch correction tools offers visualization of the sung pitch shown as a graph of semitone notes as a function of time. The instances where the singers sang off-pitch can be selected and are then manually adjusted slightly, so the result of these edits offer refined in-tune singing instances. Today, almost every CCM has vocal tracks processed with pitch correction tools. However, it should be noted that heavy processing of pitch correction tools can make the performance sound unnatural. Although there are genres of music that intentionally use these artificially pitch-processed voices, pitch correction tools are still dependent on the skillset of the singer to approximately sing the target pitch within an error range of one semitone for a relatively natural result.

The singing voices that are manipulated during the mixing process pass through the multiple speaker systems such as a pair of high-resolution dynamic loudspeakers or some consumer grade speakers. Each speaker called a driver is responsible for projecting a certain bandwidth of frequencies respectively, since it may not accurately reproduce the original sound [37]. Usually, the system is subdivided into 2 drivers with one being the woofer (responsible for the reproduction of the frequencies in the range of 20 Hz to 2 kHz), and a tweeter (responsible for the reproduction of the frequencies in the range of 2 kHz to 20 kHz). The combined two drivers replicate a more accurate sound. A third mid-range driver producing a frequency range of 500 Hz to 4 kHz is sometimes added with the woofer and the tweeter, so that the woofer and the tweeter can be more regionalized in projecting a narrower range of frequencies to further recreate an even more accurate audio signal [37]. Because of the physical natures of the speakers, there are certain points in the frequency spectrum where one driver starts to fail in accurately, recreating certain frequencies, and another driver starts to take over. This is described as an audio crossover, and because of such phenomena, the speaker cannot be completely neutral in its frequency response.

When the original audio passes through the recording phase, it is also noteworthy that all outboard audio gears can make subtle changes and distortions to the original sound. The maximal outputs of the gears should be manipulated or controlled by the audio engineers to no more than a certain loudness level. However, when the maximal output levels are beyond a certain level, the peaks of the output are clipped by various components within the audio equipment such as tubes, transistors, capacitors, integrated circuits, or tape, called peak clip-

ping. The level of distortion varies as a function of the voltage of the signal. Because signals with higher intensity are distorted further, they cause more change to the original sound. Therefore, the engineers should monitor the distortions from all instruments and seek ways to reduce unwanted distortions if it is found.

Techniques for Manipulating Intensity of the Singing Voice in Recording Studios

While there are ways to increase the intensity of the voice physiologically by strategic uses of the subglottic lung pressure, the adduction of the vocal folds, and the changes in shape of the vocal tract, audio engineers use some techniques to limit the maximal output intensity of the singing voice digitally. One of them is raising the gain level until an audible peak clipping is heard, as described above. Peak clipping is a property of a linear amplification. The maximal outputs beyond a certain loudness level are cut or saturated and listeners with normal hearing usually cannot tolerate such sounds. One of the problems induced by this technique is distortion. To reduce the distortions from peak clipping, another technique called the dynamic audio compression methods have been used in song productions. Compression modulates the amount of gain when the input or output levels increase. The gain of the acoustic or electrical systems increases the soft sounds and gradually decreases the greater intensity of sounds. This is a property of a non-linear amplification. Compression has been used to optimize the volume range of instrument to the dynamic range of listeners with normal hearing. The dynamic range is the difference between the hearing threshold and the uncomfortable loudness level. Listeners with normal hearing have a wide dynamic range of approximately 100 dB, while listeners with sensorineural hearing loss have a narrower dynamic range because of the loudness recruitment referring to an abnormal rapid growth of loudness. The audio engineers use compression to limit the loud sounds to a comfortable loudness level while making quiet sounds louder across the whole audio track [38]. Therefore, the use of compression can result in a more evenly distributed amplitude. Digitally sampled instruments and sounds are usually compressed to some extent to fit the needs of the listeners with normal hearing.

However, in a digital processing era where music can be produced exclusively with VSTs and virtual instruments, the quiet section of the singing voice can be masked by the already consistent amplitude of other melodic instruments, while the loudest part of the vocals may stand out excessively from other instruments. This produces an uneven volume balance in the mixing process. In a recording environment, the human voice is a melodic instrument with one of the largest dynamic ranges, so the singing voices are usually compressed to an extreme degree in order to be heard consistently throughout the whole song (see Figure 2).

During the mastering process, the engineers must choose to use compression on the final stereo track to make it generally louder or keep the dynamic levels in a more natural state, which is a trade-off that they must consider. There has been an ongoing trend where the engineers consistently aim to attain a loud volume throughout the whole song to make it stimulating to the listener's ears. This is referred to as the "Loudness War" [39] and in most genres in CCM, all the instruments including the vocals are overly compressed to a point where it is difficult to discriminate between the loudest part from the softest part, since the whole audio data is virtually "squared out", as seen in Figure 3. Although some of the current audio streaming platforms are now beginning to normalize the volume of their distributed songs [40], mastering engineers still practice this over-compression method to satisfy the needs of the listeners or producers of the song. This may be hazardous to novice voice students when trying to sing with a matched intensity to a singing voice that has been over-compressed.

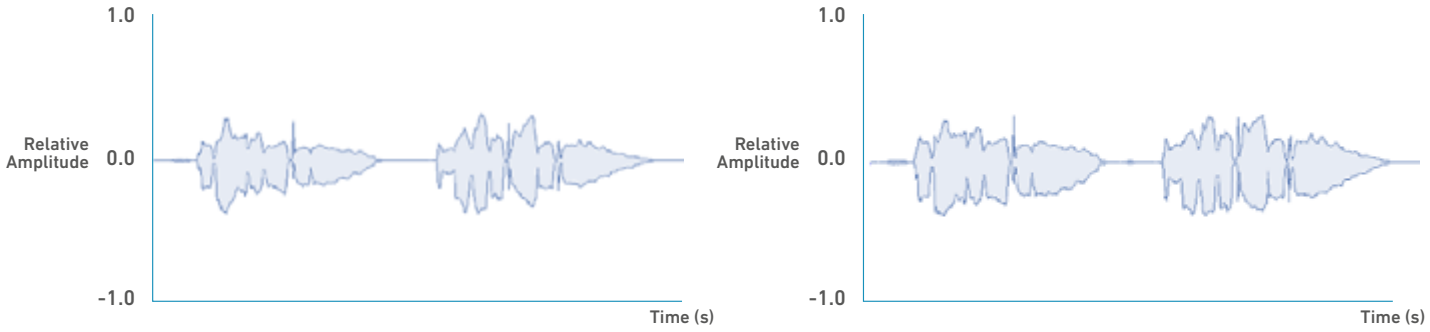

Figure 2. Uncompressed audio waveform (left), and compressed audio waveform (right).

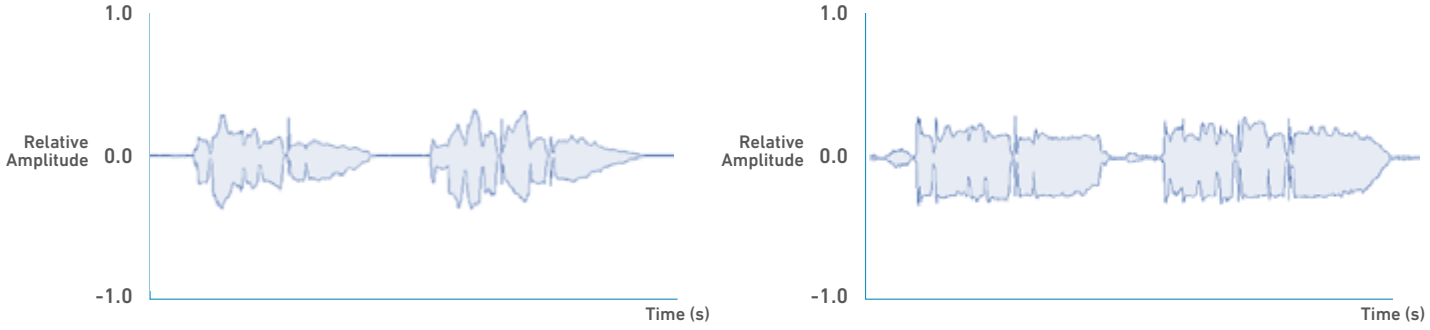

Figure 3. Original vocal waveform (left), and over-compressed vocal waveform (right).

Conclusions

We live in the changing era that uses both the traditional voice training and the new voice processing methods by audio engineers to achieve a better quality in the singing voice. The singing voices that are heard in current CCM are processed with pitch correction tools to tune the small inaccuracies in pitch during a performance, and the quality of the voice is changed using equalizers and dynamic audio compressors. Meanwhile, the major focus of voice training in physiological aspects has been to improve the voice timbre and reinforce the intensity of the singing voice while simultaneously pertaining to vocal hygiene.

Although one can argue that voice training is now relatively redundant due to the advent of digital processing, there are skillsets that singers must possess that the current audio engineers cannot replace through their techniques. In the scope of engineering and practice, skilled singers must have refined control over their articulators to manipulate the shape of their resonators. Modifying the vowel by manipulating the shape of the vocal tract can help singers maintain a uniform timbre throughout the passaggios and induce a resonant, powerful voice all throughout the singer's range. This can cause the formants to change in frequency, resulting in changes of the voice timbre that digital engineers cannot yet replicate. Furthermore, singers must train their internal vocal mechanisms to control their loudness while maintaining good vocal hygiene. One example to achieve this is the messa di voce technique, a technique referring to singing in a single pitch while slowly intensifying and decreasing the

loudness of the voice. An important aspect of this technique is to maintain a natural sonority in the voice while singing without excessive effort. Depending heavily on the subglottal lung pressure during these exercises may lead to singing in a pressed voice and may cause vocal hygiene problems among novice voice professionals. Adjusting the shape of the vocal tract and reconfiguring the glottal width during singing may be viable alternatives to making the singing voice louder [27-29, 31]. Singers must strive to acquire the proper techniques to freely change their voice while simultaneously maintaining their vocal health.

References

1. Nielsen MRC Soundscan, Billboard. MRC Data Year-End Report - Billboard [Internet]. [cited 2021 Nov 19]. Available from: https://static.billboard.com/files/2021/01/MRC_Billboard_YEAR_END_2020_US-Final201.8.21-1610124809.pdf
2. Lanus L. Is Classical Music not popular anymore? Pulsechamber Music [Internet]. 2018 [cited 2021 Nov 19]. Available from: https://pulsechambermusic.com/is-classical-music-not-popular-anymore/
3. RIAA. U.S. sales database [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://www.riaa.com/u-s-sales-database/
4. Wikipedia contributors. Sound recording and reproduction. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Sound_recording_and_reproduction#1950s_to_1980s
5. Wikipedia contributors. Steinberg Cubase. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Steinberg_Cubase
6. Wikipedia contributors. Pro tools. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Pro_Tools
7. Wikipedia contributors. Singing. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Singing
8. Pershall KE, Boone DR. Supraglottic contribution to voice quality. J Voice. 1987;1(2):186–90. doi: https://doi.org/10.1016/s0892-1997(87)80044-9
9. Painter C. The laryngeal vestibule and voice quality. Arch Otorhinolaryngol. 1986;243(5):329–37. doi: https://doi.org/10.1007/bf00460212
10. Sundberg J. Articulatory interpretation of the "singing formant". J Acoust Soc Am. 1974;55(4):838–44. doi: https://doi.org/10.1121/1.1914609
11. Hunter EJ, Titze IR. Overlap of hearing and voicing ranges in singing [Internet]. Journal of singing: the official journal of the National Association of Teachers of Singing. U.S. National Library of Medicine; 2005;61(4):387-392 [cited 2021 Nov 19]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19844607
12. Shaw EAG. The external ear. In: Keidel WD, Neff WD, editors. Handbook of Sensory Physiology. New York City, New York: Springer-Verlag; 1974. p.455–90.
13. Titze IR, Story BH. Acoustic interactions of the voice source with the lower vocal tract. J Acoust Soc Am. 1997;101(4):2234–43. doi: https://doi.org/10.1121/1.418246

14. Titze IR. Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings. J Speech Lang Hear Res. 2006;49(2):448–59. doi: https://doi.org/10.1044/1092-4388(2006/035)

15. Samlan RA, Kreiman J. Perceptual consequences of changes in epilaryngeal area and shape. J Acoust Soc Am. 2013;136(5):2798–806. doi: https://doi.org/10.1121/1.4799525

16. Titze IR, Abbott KV. Vocology: The science and practice of Voice Habilitation. Salt Lake City, Utah: National Center for Voice and Speech; 2012.

17. Titze I. Acoustic interpretation of Resonant Voice. J Voice. 2001;15(4):519–28. doi: https://doi.org/10.1016/s0892-1997(01)00052-2

18. Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in Resonant Voice. J Voice. 1998;12(3):315–27. doi: https://doi.org/10.1016/s0892-1997(98)80021-0

19. Yanagisawa E, Estill J, Kmucha ST, Leder SB. The contribution of aryepiglottic constriction to "ringing" voice quality—a videolaryngoscopic study with acoustic analysis. J Voice. 1989;3(4):342–50. doi: https://doi.org/10.1016/s0892-1997(89)80057-8

20. Titze IR, Palaparthi A, Cox K, Stark A, Maxfield L, Manternach B. Vocalization with semi-occluded airways is favorable for Optimizing Sound Production. PLoS Comput. Biol. 2021;17(3): e1008744. doi: https://doi.org/10.1371/journal.pcbi.1008744

21. Titze IR, Laukkanen A-M. Can vocal economy in phonation be increased with an artificially lengthened vocal tract? A Computer Modeling Study. Logoped. Phoniatr. Vocol. 2007;32(4):147–56. doi: https://doi.org/10.1080/14015430701439765

22. Bele IV. Artificially lengthened and constricted vocal tract in vocal training methods. Logoped Phoniatr Vocol. 2005;30(1):34–40. doi: https://doi.org/10.1080/14015430510006677

23. Laukkanen A-M, Horáček J, Krupa P, Švec JG. The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results. Biomed Signal Process Control. 2012;7(1):50–7. doi: https://doi.org/10.1016/j.bspc.2011.02.004

24. Titze IR. Principles of Voice production. Iowa City, Iowa: National Center for Voice and Speech; 2000.

25. Callaghan J. Singing and voice science. San Diego, California: Singular Publishing Group; 2000.

26. Titze IR. Formant Frequency Shifts for Classical and Theater Belt Vowel Modification [Internet]. National Association of Teachers of Singing. [cited 2021 Nov 19]. Available from: https://www.nats.org/cgi/page.cgi/_article.html/Journal_of_Singing/Formant_Frequency_Shifts_for_Classical_and_Theater_Belt_Vowel_Modification_2011_Jan_Feb

27. Titze IR, Sundberg J. Vocal intensity in speakers and singers. J Acoust Soc Am. 1991;90(4):2936–46. doi: https://doi.org/10.1121/1.402160

28. Sundberg J. What's so special about singers? J Voice. 1990;4(2):107–19. doi: https://doi.org/10.1016/s0892-1997(05)80135-3

29. Titze IR. Simulation of vocal loudness regulation with lung pressure, vocal fold adduction, and source-airway interaction. J Voice. Forthcoming 2021. doi: https://doi.org/10.1016/j.jvoice.2020.11.030

30. Titze IR. Regulation of laryngeal resistance and maximum power transfer with semi-occluded airway vocalization. J Acoust Soc Am. 2021;149(6):4106–18. doi: https://doi.org/10.1121/10.0005124

31. Titze IR, Worley AS. Modeling source-filter interaction in belting and high-pitched operatic male singing. J Acoust Soc Am. 2009;126(3):1530–40. doi: https://doi.org/10.1121/1.3160296

32. Decker TN. Instrumentation: An introduction for students in the speech and hearing sciences. New York: Longman; 1990.

33. Fox A. What is microphone frequency response?. My New Microphone [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://mynewmicrophone.com/frequency-response/

34. Lass NJ, Woodford CM. Hearing science fundamentals. St. Louis, Mississippi: Plural Publishing, Inc.; 2007.

35. Owinski B. Mixing engineer's Handbook. Bobby Owsinski [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://bobbyowsinski.com/mixing-engineers-handbook/

36. Hobbs J. How to EQ vocals professionally: The Easy 6 Step Method. LedgerNote [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://ledgernote.com/columns/mixing-mastering/how-to-eq-vocals/

37. Gavin B. What are woofers, mid-range speakers, and tweeters?. How. How-To Geek [Internet]. 2018 [cited 2021 Nov 19]. Available from: https://www.howtogeek.com/354985/what-are-woofers-mid-range-speakers-and-tweeters/

38. Wikipedia contributors. Dynamic Range Compression. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Dynamic_range_compression

39. Wikipedia contributors. Loudness War. Wikipedia. Wikimedia Foundation [Internet]. 2021 [cited 2021 Nov 19]. Available from: https://en.wikipedia.org/wiki/Loudness_war

40. Stewart I. Mastering for streaming platforms: Normalization, LUFS, and loudness myths demystified. iZotope [Internet]. 2014 [cited 2021 Nov 19]. Available from: https://www.izotope.com/en/learn/mastering-for-streaming-platforms.html