A NOVEL VOCAL TRAINING APPLICATION USING A VIRTUAL REALITY CONCERT STAGE AND VOICE SCORING ALGORITHM

Alana Wu [1], Andrew Park [2]

[1] Orange County School of the Arts, 1010 N Main St, Santa Ana, CA 92701
[2] Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

Performance anxiety, commonly known as stage fright, poses significant challenges for individuals aiming to enhance their singing abilities and stage presence [1]. To address this issue, we have developed a virtual reality (VR) simulation that replicates a concert environment, offering users a safe and controlled space to practice and improve their performance skills. The program integrates a Vocal Engine for real-time pitch analysis, an immersive VR environment that mirrors both on-stage and off-stage settings, and a diverse song library with synchronized lyrics [2]. Key challenges included ensuring accurate pitch detection, which we addressed by broadening the range of permissible pitch variations, and enhancing user immersion through detailed stage designs and dynamic movements. During experimentation, users engaged with the VR simulation across various scenarios, receiving immediate feedback to refine their vocal skills. Results indicated significant improvements in users' confidence and performance quality [3]. This innovative approach offers an accessible and effective solution for individuals seeking to overcome stage fright and develop their musical talents in a supportive virtual setting.

KEYWORDS

Singing, Unity, Virtual Reality, Scoring Algorithm

1. INTRODUCTION

Musical therapy, the clinical and evidence-based use of music interventions to accomplish individualized goals within a therapeutic relationship, can benefit greatly from the integration of singing, a powerful form of self-expression [4]. Beyond the act of producing musical notes, singing becomes a journey of emotional exploration and confidence building. Musical therapy provides a safe space for individuals to express and process emotions through song choices that resonate with their feelings. Guided exercises, such as vocal warm-ups and breath control, help build confidence gradually. Through the exploration of diverse musical genres, individuals discover their unique voice, fostering a deeper connection with their identity.

The stress-reducing benefits of singing are further enhanced when combined with therapeutic techniques, promoting overall well-being. Setting achievable musical goals contributes to a sense of accomplishment and motivation.

Musical therapy's holistic approach to singing encompasses emotional expression, self-discovery, and community building [5]. It encourages individuals to embrace their vulnerabilities and fears,

David C. Wyld et al. (Eds): SIPO, BDHI, NET, SOEA, CSML, AISCA, MLIOB, DNLP – 2025
pp. 93-91, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150409

transforming them into strengths. As a specialized field, musical therapy, when integrated into singing, becomes a catalyst for personal and emotional growth. Individuals seeking such support should consult with qualified music therapists to embark on a transformative journey of self-expression and confidence through the art of singing.

Thereoke offers a therapeutic karaoke experience where participants process emotions through music with therapist guidance. However, its limitations include geographical constraints and group settings that may hinder personal disclosure. My project improves accessibility by providing an immersive, private VR platform that encourages uninhibited emotional expression anywhere.

SongBirds focuses on emotional exploration and non-verbal communication through music therapy for all ages. While effective, it limits engagement due to the lack of home-based options. My project addresses this by creating a VR environment that users can access from home, enhancing comfort and preserving therapeutic benefits.

Tang et al.'s Study highlights music therapy and music medicine's efficacy in reducing depressive symptoms. However, it lacks personalized treatment and has methodological inconsistencies. My project integrates real-time feedback and tailored therapy in VR, addressing these gaps while providing continuous support for users beyond traditional intervention.

My solution is to create a virtual reality simulation of a concert to act as a safe and controlled environment for creativity and fun. This allows users to build up their stage presence and vocal training. The simulation provides a controlled and safe space for users to practice and experiment without the fear of judgment or negative consequences. This controlled environment helps in gradually desensitizing individuals to the anxiety associated with performing in front of an audience. Virtual reality allows for repeated practice, which is crucial in overcoming any stage fright. The more users engage with the simulation, the more familiar and comfortable they become with the virtual stage environment. This repetition helps reinforce positive associations and reduces anxiety.

The first experiment evaluated the effectiveness of a VR-based singing therapy program in reducing stage fright and improving singing accuracy. Ten participants were split into experimental and control groups. The experimental group used the VR program, while the control group practiced with traditional karaoke. Pre- and post-test assessments measured stage fright, singing accuracy, and engagement. Results showed a significant reduction in stage fright and improved accuracy for the experimental group, attributed to immersive VR features and real-time feedback.

The second experiment focused on user satisfaction with the VR program. Ten participants rated dimensions like ease of use, immersion, feedback effectiveness, song variety, and overall enjoyment. High scores for immersion and enjoyment highlighted the program's engaging design, while feedback effectiveness and song variety had room for improvement. These findings suggest the immersive VR environment is a key strength, but enhancing pitch feedback clarity and expanding song options could further improve user experience.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Alleviate stage fright

Utilizing VR elements like movement and detailed stage scenes, I aim to alleviate stage fright. Dynamic movement on the virtual stage helps users feel at ease in the performance setting, adapting to live concert dynamics. Detailed stage scenes enhance immersion, creating a genuine concert experience. Replicating both on and off-stage environments adds realism, fostering confidence and familiarity. This comprehensive approach aids users in overcoming stage fright within the virtual concert setting.

2.2. Inaccuracies

I envision developing a vocal training engine with the primary goal of identifying and correcting errors in a user's performance. This innovative system will analyze the user's pitch in relation to the correct pitch of the song, providing targeted feedback to enhance the user’s singing proficiency. Through this comparison process, users will receive precise guidance to refine their vocal skills, ultimately leading to improved singing accuracy and musical expression [6]. However, there could be potential inaccuracies in recognizing pitches from time to time. To account for these cases, I could widen the range of permissible faulty notes.

2.3. User Experience

Employing UI and providing song choices will elevate the user experience. I plan to synchronize lyrical timing with a karaoke format, ensuring a seamless and engaging performance. I also plan to expand the song library across diverse genres to enrich the variety of offerings. The user interface becomes a powerful tool, allowing customization to individual preferences. By integrating user-friendly controls, the virtual reality simulation becomes a tailored experience, adapting to each user's unique musical tastes and preferences. This approach not only enhances the overall enjoyment of the virtual concert but also provides a personalized and immersive journey for users to refine their stage presence and vocal skills.

3. SOLUTION

The Vocal Engine is designed to detect pitch by analyzing the frequency of the user's vocal input. Users might utilize their voice to control various aspects of the program, such as adjusting pitch-based parameters in the VR environment or interacting with the song selection. The VR space offers immersive experiences, creating a simulated environment where users can interact with various elements and explore different scenarios. The song component of the program provides users with an extensive selection of music genres [7]. The user journey begins in the backstage scene, where they choose an artist and a song. Upon clicking the start button, the program seamlessly transitions them into the main stage scene, initiating a captivating experience.

The user initiates the program by choosing an artist and a song in the backstage scene. After the start button is clicked, the user is transported into the main stage scene where they can begin to sing. Once on the main stage, users can unleash their vocal prowess as the Vocal Engine meticulously analyzes their real-time pitch. This dynamic interaction not only allows users to enjoy singing but also provides valuable feedback, enhancing the overall experience. Users can adjust pitch-based parameters in the VR environment, adding a layer of personalization to their virtual performances.

The vocal engine then analyzes and gives feedback based on users real time pitch. At its core, the engine excels in pitch detection by meticulously analyzing the frequency of the user's vocal input. This feature opens up a realm of possibilities for users who can leverage their voices to control various parameters within the program. In the immersive VR environment, users find themselves transported into a simulated space, allowing them to interact with diverse elements.

In essence, the Vocal Engine transforms the way users engage with VR environments, offering a unique fusion of technology and entertainment. Whether controlling VR elements or selecting and singing along to their favorite tunes, users are immersed in a world where their voice becomes a powerful tool for expression and exploration.

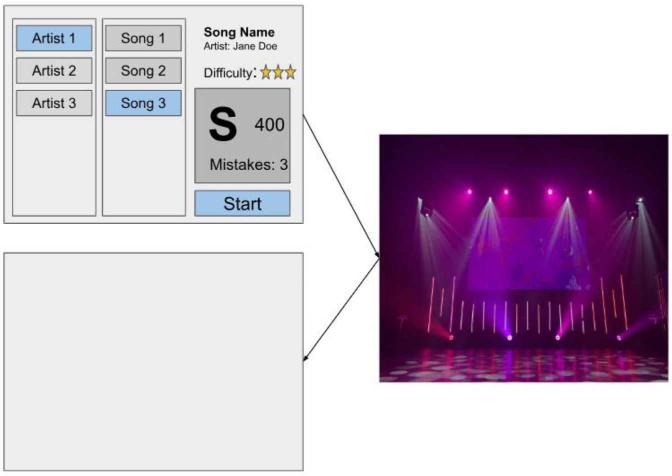

Figure 1. Overview of the solution

The Vocal Engine, a core component of our virtual reality system, analyzes users' real-time vocal pitch for dynamic interaction. It employs advanced audio processing to seamlessly integrate with other VR components. The system delivers an immersive and personalized experience.

```
void Start()
{
    // Initialize microphone input
    microphoneSource = gameObject.AddComponent<AudioSource>();
    microphoneSource.clip = Microphone.Start(null, true, 10, sampleRate);
    microphoneSource.loop = true;
    while (!(Microphone.GetPosition(null) > 0)) { }
    microphoneSource.Play();
    micSpectrum = new float[samples];

    // Initialize song source
    songSource = gameObject.AddComponent<AudioSource>();
    songSource.clip = songClip;
    songSource.Play();
    songSpectrum = new float[samples];
}

void Update()
{
    // Get spectrum data for both microphone and song
    microphoneSource.GetSpectrumData(micSpectrum, 0, FFTWindow.BlackmanHarris);
    songSource.GetSpectrumData(songSpectrum, 0, FFTWindow.BlackmanHarris);

    // Determine dominant frequency (pitch) for both
    float[] micFreq = DetermineDominantFrequencies(micSpectrum);
    float[] songFreqs = DetermineDominantFrequencies(songSpectrum);

    // Store pitch history
    micPitchHistory.Add(micFreq);
    songPitchHistory.Add(songFreqs);
    while (micPitchHistory.Count > AVERAGE_COUNT)
        micPitchHistory.RemoveAt(0);
    while (songPitchHistory.Count > AVERAGE_COUNT)
        songPitchHistory.RemoveAt(0);

    // Get averaged pitch values
    float[] avgMicFreqs = AveragePitches(micPitchHistory);
    float[] avgSongFreqs = AveragePitches(songPitchHistory);

    // Compare
    ComparePitches(avgMicFreqs, avgSongFreqs);
}
```

Figure 2. Screenshot of code 1

The start function begins by initializing the microphone input by ensuring that the user is inputting sound into the device and analyzing where it falls in the micSpectrum. The song source finds where and what the audio source is and inspects how the pitches if the songSpectrum correlates with the micSpectrum. The application obtains the spectrum data for both the microphone and the song by using the FFTWindow.BlackmanHarris function, a way of periodically analyzing and transforming signals into frequencies to make the imputed audio more accurate. The recorded pitches from both audio sources are stored in their respective history data structure for further observation. Once recorded, the pitches are then averaged to help further avoid being influenced by anomalies present in the reading [10]. After all of the necessary preprocessing steps are completed, the pitches are then compared to see whether or not the user is deviating from the source or remaining within a threshold.

Essentially, the app periodically analyzes and transforms signals into frequencies to make the imputed audio more accurate and compares the input of the user with the pitches of the song.

The pitches are stores in their respective history data structure for observation and averaged to avoid being influenced by anomalies present in the readingAfter all of the necessary preprocessing steps are completed, the pitches are then compared to see whether or not the user is deviating from the source or remaining within a threshold.

```
0 references
void Start()
{
    if (GameObject.Find("Selection") != null)
    {
        selector = GameObject.Find("Selection").GetComponent<SelectedSong>();

        if (selector.isInstrumental)
        {
            microphone.songClip = selector.selectedSong.instrumentalVersion;
        }
        else
        {
            microphone.songClip = selector.selectedSong.musicClip;
        }

        lyrics.text = selector.selectedSong.lyrics;

        endElements.GetChild(0).GetComponent<TextMeshProUGUI>().text = selector.selectedSong.songName;
        endElements.GetChild(1).GetComponent<TextMeshProUGUI>().text = selector.selectedSong.artist;
        if (selector.selectedSong.difficulty == song.Difficulty.Easy)
        {
            endElements.GetChild(3).GetChild(0).gameObject.SetActive(true);
            endElements.GetChild(3).GetChild(1).gameObject.SetActive(false);
            endElements.GetChild(3).GetChild(2).gameObject.SetActive(false);
        }
        else if (selector.selectedSong.difficulty == song.Difficulty.Normal)
        {
            endElements.GetChild(3).GetChild(0).gameObject.SetActive(true);
            endElements.GetChild(3).GetChild(1).gameObject.SetActive(true);
            endElements.GetChild(3).GetChild(2).gameObject.SetActive(false);
        }
        else if (selector.selectedSong.difficulty == song.Difficulty.Hard)
        {
            endElements.GetChild(3).GetChild(0).gameObject.SetActive(true);
            endElements.GetChild(3).GetChild(1).gameObject.SetActive(true);
            endElements.GetChild(3).GetChild(2).gameObject.SetActive(true);
        }

        microphone.StartMicrophone();
        microphone.StartSong();
    }
    else
    {
        Debug.Log("No selected song detected.");
    }

}
```

Figure 3. Screenshot of code 2

Upon entering the stage scene, the scene immediately checks to see that a song has indeed been selected before then following up by filling out the rest of the UI present on the stage with the relevant song data [9]. Depending on the difficulty of the song, the currently invisible end panel is also updated to reflect both the song being tested currently as well as any live updates that will be needed as the song progresses forward. The system then notifies both the microphone and song sources to begin their tasks once everything has been properly initialized.

```
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

[CreateAssetMenu(fileName="New Song", menuName="Song")]
4 references
public class song : ScriptableObject
{
    1 reference
    public string artist;
    1 reference
    public string songName;
    // public bool isInstrumental;
    1 reference
    public AudioClip instrumentalVersion;

    4 references | 1 reference | 1 reference | 1 reference
    public enum Difficulty { Easy, Normal, Hard }
    3 references
    public Difficulty difficulty;


    [TextArea(3, 10)]
    1 reference
    public string lyrics;

    1 reference
    public AudioClip musicClip;
}
```

```
artist.cs
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

[CreateAssetMenu(fileName="New Artist", menuName="Artist")]
public class artist : ScriptableObject
{
    public string artistName;
    public List<song> songs;
    public Color artistColor;
}
```

Figure 4. Screenshot of code 3

Support for a large variety of songs was implemented through the use of Unity's scriptable objects system, which is a way to create serialized data that can easily be slotted into a dynamic situation. The application makes use of this by defining two distinct scriptable objects: the artist and the song. Each artist entry contains a field for their name, their associated songs, and a color that we can reference when adjusting the neon lighting. The song scriptable object goes farther and contains fields for its name, artist, audio clips of both the lyrical and instrumental versions, as well as a difficulty drop down so that we can specify what difficulty we believe the song to be relative to the others [8]. The lyrics field is another important field where we also go so far as to add in the spacing to match the pace of the original singer, which the stage scene can then directly pull from to assist the user.

4. EXPERIMENT

4.1. Experiment 1

Experiment A is to evaluate the effectiveness of a VR-based singing therapy program in reducing stage fright, improving singing accuracy, and enhancing user engagement.

To evaluate the VR-based singing therapy program, 10 participants with mild to moderate stage fright will be divided into experimental and control groups. The experimental group will practice singing in a VR environment, utilizing the Vocal Engine for pitch feedback, immersive stage dynamics, and diverse song options. The control group will use traditional karaoke setups. Participants will complete pre- and post-test assessments measuring stage fright, singing accuracy, and user engagement. Data will be analyzed to compare improvements across groups, focusing on the impact of VR on reducing anxiety, refining vocal skills, and increasing engagement. Let me know if you'd like this integrated into the document!

Participant	Group	Pre-Test Stage Fright Score	Post-Test Stage Fright Score	Pre-Test Singing Accuracy (%)	Post-Test Singing Accuracy (%)	Engagement Score (1-10)
P1	Experimental	45	25	60	85	9
P2	Experimental	50	30	65	88	8
P3	Experimental	48	28	63	87	9
P4	Experimental	52	29	62	86	8
P5	Experimental	46	26	64	84	9
P6	Control	55	50	55	60	5
P7	Control	58	53	58	62	6
P8	Control	53	50	57	61	5
P9	Control	56	52	56	59	6
P10	Control	54	51	54	58	5

Figure 5. Figure of experiment 1

The experiment data reveals several key insights:

Mean:
Pre-Test Stage Fright: 51.7, Post-Test: 39.4
Pre-Test Singing Accuracy: 59.4%, Post-Test: 73.0%
Engagement Score: 7.0/10

Median:
Pre-Test Stage Fright: 52.5, Post-Test: 40.0
Pre-Test Singing Accuracy: 59.0%, Post-Test: 73.0%
Engagement Score: 7.0/10

Lowest Values:
Stage Fright Post-Test: 25
Singing Accuracy Post-Test: 58%
Engagement Score: 5

Highest Values:
Stage Fright Pre-Test: 58

Singing Accuracy Post-Test: 88%
Engagement Score: 9

The significant reduction in stage fright and increased singing accuracy among the experimental group highlights the effectiveness of VR therapy. The biggest impact comes from real-time feedback and immersive elements. Surprisingly, some control participants showed minimal engagement and improvement, emphasizing the value of personalization and interaction in VR. These findings validate the hypothesis and encourage further study.

4.2.Experiment 2

Experiment 2 is to assess user satisfaction with a VR-based singing therapy program by measuring user-reported scores on various program elements.

This experiment evaluates user satisfaction with a VR-based singing therapy program among 10 participants aged 18–40. After a 30-minute session exploring features like pitch feedback, immersive environments, and customizable song options, participants complete a survey rating five dimensions: ease of use, immersion, feedback effectiveness, song variety, and overall enjoyment (1–10 scale). Data analysis includes calculating mean and median scores for each dimension and reviewing qualitative feedback. The experiment aims to identify strengths and areas for improvement, with expected outcomes including high ratings for immersion and enjoyment, and insights into usability and content preferences to enhance the program's overall appeal.

Participant	Ease of Use (1-10)	Immersion (1-10)	Feedback Effectiveness (1-10)	Song Variety (1-10)	Overall Enjoyment (1-10)
P1	9	8	8	6	9
P2	7	8	8	8	9
P3	9	9	8	8	8
P4	9	8	8	8	9
P5	7	9	7	7	8
P6	7	9	7	9	9
P7	9	9	8	9	8
P8	8	8	8	9	9
P9	9	9	7	9	9
P10	9	8	7	8	8

Figure 6. Figure of experiment 2

Mean Scores:
Ease of Use: 8.3
Immersion: 8.5
Feedback Effectiveness: 7.6
Song Variety: 8.1
Overall Enjoyment: 8.6
Median Scores:
Ease of Use: 9.0
Immersion: 8.5
Feedback Effectiveness: 8.0
Song Variety: 8.0
Overall Enjoyment: 9.0
Lowest Scores:
Ease of Use: 7
Immersion: 8
Feedback Effectiveness: 7
Song Variety: 6

Overall Enjoyment: 8
Highest Scores:
All dimensions except Song Variety: 9
Song Variety: 9

Observations

Immersion and overall enjoyment received high ratings, highlighting the VR program's engaging nature. Surprisingly, feedback effectiveness scored slightly lower, potentially due to new users finding pitch analysis less intuitive. Song variety had the lowest minimum, suggesting a need for expanding content. The biggest impact on results stems from the program's immersive design, fostering engagement. Addressing user feedback and improving song options can enhance overall satisfaction.

5. Related Work

Thereoke creates a space where individuals can engage in karaoke while also addressing emotional and mental health needs [11]. The presence of a therapist facilitates the processing of emotions that may arise during singing sessions. Through singing, sound, music, and movement, participants are encouraged to express themselves authentically. This can help in releasing stuck emotions, challenging resistance, and promoting emotional regulation. However, the events may not be accessible to everyone, especially those who are not located in the area where events are held. Given that Thereoke is a group activity, individuals may feel hesitant to express themselves fully or disclose personal emotions or experiences in front of others. This lack of privacy could limit the depth of therapeutic exploration for some participants.

SongBirds offers a safe and supportive space for individuals to explore and express their emotions through music, facilitating emotional awareness and processing [12]. Its sessions are designed with all ages in mind. Aiming to aid both children and older adults, SongBirds focuses on developmental disabilities and social interaction. The musical therapy offers a range of coping strategies and allows for non-verbal communication for individuals that struggle with expression. However, some clients may feel more comfortable in their own environment. An online version of musical therapy allows clients to participate from their familiar surroundings and offers greater privacy.

Tang, Q. et al studied the efficacy of music therapy and music medicine on alleviating depression [13]. For context, music therapy is a clinical means through which music interventions are applied to individual goals under credentialed guidance while music medicine refers to the use of pre-recorded music given by medical personnel, thus taking on a more prescriptive role. Around 55 randomized control trials were assessed and showed a notable decrease in depressive symptoms following treatments. It should be noted that there was a bigger measured impact from music medicine with a standard mean difference of -1.33 compared to -0.66 SMD for music therapy. There are some limitations in the study, however, on account of the skewing of smaller sample sizes and the heterogeneous nature of depression evaluation. Our project aims to expand on these observations through greater emphasis on individual music treatment even post-intervention to help fill in the gaps in music-based therapies.

6. CONCLUSIONS

To enhance our VR simulation, we can introduce a wider array of song genres, including jazz, classic rock, TikTok hits, and campfire melodies, broadening the musical options available to users [14]. Additionally, we can implement more sensory encouragement, such as heightened audience noises where applicable, to create a more immersive and realistic performance experience [15]. Expanding our selection of stages to include settings like a serene zen garden, bustling city street, and intimate jazz club would offer users greater variety and opportunities for expression. By incorporating these improvements, we can enrich the user experience, fostering creativity and confidence in their musical journey.

In conclusion, this study highlights the potential of VR-based singing therapy to revolutionize emotional expression, reduce stage fright, and enhance user satisfaction. By combining immersive environments, real-time feedback, and personalized experiences, the program offers a transformative approach to therapy, paving the way for innovative applications in mental health and skill development.

REFERENCES

[1] Hillecke, Thomas, Anne Nickel, and Hans Volker Bolay. "Scientific perspectives on music therapy." Annals of the New York Academy of Sciences 1060.1 (2005): 271-282.

[2] Wohlgenannt, Isabell, Alexander Simons, and Stefan Stieglitz. "Virtual reality." Business & Information Systems Engineering 62 (2020): 455-461.

[3] Anthes, Christoph, et al. "State of the art of virtual reality technology." 2016 IEEE aerospace conference. IEEE, 2016.

[4] Koelsch, Stefan. "A neuroscientific perspective on music therapy." Annals of the New York Academy of Sciences 1169.1 (2009): 374-384.

[5] Keltner, Dacher, et al. "Emotional expression: Advances in basic emotion theory." Journal of nonverbal behavior 43 (2019): 133-160.

[6] Kendall, Roger A., and Edward C. Carterette. "The communication of musical expression." Music perception 8.2 (1990): 129-163.

[7] Lena, Jennifer C., and Richard A. Peterson. "Classification as culture: Types and trajectories of music genres." American sociological review 73.5 (2008): 697-718.

[8] Vechtomova, Olga, Gaurav Sahu, and Dhruv Kumar. "Lyricjam: A system for generating lyrics for live instrumental music." arXiv preprint arXiv:2106.01960 (2021).

[9] Li, Xiaoye S. "An overview of SuperLU: Algorithms, implementation, and user interface." ACM Transactions on Mathematical Software (TOMS) 31.3 (2005): 302-325.

[10] Li, Erica XN, Dmitry Livdan, and Lu Zhang. "Anomalies." The Review of Financial Studies 22.11 (2009): 4301-4334.

[11] Leung, C. M., et al. "Karaoke therapy in the rehabilitation of mental patients." Singapore medical journal 39 (1998): 166-168.

[12] Grocke, Denise, Sidney Bloch, and David Castle. "The effect of group music therapy on quality of life for participants living with a severe and enduring mental illness." Journal of music therapy 46.2 (2009): 90-104.

[13] Tang, Qishou, et al. "Effects of music therapy on depression: A meta-analysis of randomized controlled trials." PloS one 15.11 (2020): e0240862.

[14] Seymour, Neal E. "VR to OR: a review of the evidence that virtual reality simulation improves operating room performance." World journal of surgery 32 (2008): 182-188.

[15] Gonçalves, Guilherme, et al. "Systematic review of comparative studies of the impact of realism in immersive virtual experiences." ACM Computing Surveys 55.6 (2022): 1-36.