ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9747033

SLEEPGAN: TOWARDS PERSONALIZED SLEEP THERAPY MUSIC

Jing Yang[⋆∗] *Chulhong Min*[†] *Akhil Mathur*[†] *Fahim Kawsar*[†]

[†] Nokia Bell Labs, Cambridge, UK
[⋆] Department of Computer Science, ETH Zurich, Switzerland

ABSTRACT

Sleep deficiency and disorders are one of the most unsolved public health challenges of modern times. Music therapy is a promising approach, offering a cheap and non-invasive solution to improve sleep quality. However, the choice of therapeutic sleep music is highly limited for users because such music needs to be specially chosen and made by sleep therapists. It could potentially lead to the inefficiency of music therapy if users get bored after listening to the same set of music repeatedly. In this paper, we take the first step towards generating personalized sleep therapy music. Firstly, through an in-depth feature analysis, we investigate the importance of various musical and acoustic features of therapy music. Grounded on our findings, we design a style transfer framework called SleepGAN which induces therapeutic features into music from different genres. We show that, compared to baselines, the music generated by SleepGAN has a higher similarity to the sleep music designed by experts.

Index Terms— Music style transfer, Sleep therapy

1. INTRODUCTION

Sleep is an essential function of the human body and has a direct impact on our physical and mental well-being. Research has shown that sleep disruption and disorders have a strong causal link to major lifestyle diseases such as memory loss, obesity, diabetes, and cancer [1, 2]. Unfortunately, sleep disorders are highly prevalent in our society — studies show that nearly 40% of United Kingdom adults and roughly 50–70 million American adults experience them [3, 4]. Cognitive Behavior Therapy and pharmaceutical sleep aids are two common clinical interventions to alleviate sleep disorders. However, they are expensive and have potentially harmful side-effects [5], which has led to research in designing non-invasive, low-cost interventions to address sleep disorders.

Music therapy is a promising approach in this direction, with the goal being to expose users/patients to 'therapeutic music' prior to sleep which can enhance their sleep quality. Clinical studies have shown that sleep disorders can be mitigated using therapeutic music, due to the potential effect of such music on modulating the sympathetic nervous system activity [6], regulating the stress hormone cortisol [7], and increasing the levels of oxytocin in the body [8]. Further, commercial products such as the Bose Sleep Buds [9] have also been launched to support music therapy through earbuds.

However, music used for therapy is primarily designed or chosen by expert sleep therapists. Thus, the available therapeutic music is often highly limited in terms of scalability and accessibility. Although clinical studies have shown that a user's personal music preference could have an impact on the effectiveness of music therapy [10, 11], currently there are no effective ways to personalize therapy music for individual users. As such, users are left with a small collection of pre-designed therapy music to choose from. This, in turn, could limit mass adoption of sleep music apps, as well as potentially lead to music fatigue [12] wherein users get bored after listening to the same set of therapy music repeatedly.

The goal of this paper is to explore techniques for generating personalized sleep music. This means, users can select music from their own playlists and our algorithm can make the user-selected music more therapeutic. We envision that such a technique will enable sleep therapy apps to accept arbitrary music file, thereby allowing users to have sleep therapy with a personalized list of songs.

The key research challenges are *twofold*: Firstly, there is insufficient understanding of the relationship between various musical features and their therapeutic effect on sleep music. Although prior research has investigated the characteristics of sleep music [3, 13], the studies were done with selective user groups, or included music playlists from a small group of users. Secondly, it remains an open question how to induce therapeutic sleep features into any type of user-selected music. While [14] has looked at generating music for anxiety-reduction, no evidence has shown that music for addressing such mental problems could equally work for sleep improvement.

In a first of its kind exploration, we analyze 399 examples of sleep music taken from clinical studies and crowd-sourced Spotify sleep playlists to understand their fundamental acoustic and musical properties, and how they differ from other genres of music. By extracting a total of 34 features from each music and performing a K-means clustering on them, we learn that sleep music is primarily characterized by its bass, treble, and overall pitch profile. More specifically, spectral rolloff features are one of the most prominent discriminating features between sleep music and other types of music. Based on this in-depth feature analysis, we design a CycleGAN-based style transfer framework called SleepGAN with a customized optimization objective that aims to 'transfer' therapeutic features from sleep music into any user-specified music. Our evaluation illustrates that SleepGAN manages to generate music which — as compared to the original user-specified music — is quantitatively more similar to sleep therapy music.

In summary, our contributions include a feature analysis of sleep music to decipher what makes it different from other types of music. Thereafter, we show that the therapeutic features uncovered in our analysis could be used to design more accurate style transfer methods. As a clinical sleep study to evaluate the generated music's efficacy is out of the scope of this paper, we present an objective evaluation to show that the music generated by SleepGAN is quantitatively similar to the sleep music designed by experts. We also present a preliminary subjective study to evaluate if users prefer the music generated by SleepGAN over other baseline methods.

2. MUSICAL FEATURE ANALYSIS

In this section, we conduct feature analysis to answer the question *"what are the most discriminating musical features for sleep mu-*

[∗]This work was done at Nokia Bell Labs during the author's internship.

sic". This exploration will indicate potential strategies for developing a style transfer model that can incorporate therapeutic properties into arbitrary user-selected music. To advance an explanatory understanding of sleep music, we ground the analysis on fundamental musical features (e.g., rhythm, articulation).

In the following, we first describe the dataset used in our study, then elaborate on the feature analysis using K-means clustering.

2.1. Dataset

For the study, we create a dataset that consists of two sets of music, *sleep music* and *other music*.

Sleep music: We collected 399 pieces of sleep music (no lyrics) from the following two resources. First, we collected 26 pieces that have been proven therapeutic in clinical sleep studies [13, 15]. Second, like [16], we crowd-sourced sleep music using the popular music streaming service provider Spotify[1]. We searched for Spotify playlists that were made specifically for sleep aid. Then, to alleviate bias towards individual music preferences, we only included the top 3 voted playlists that each had more than 200K likes.

Other music: For a comparative study, we further collected 1K music and songs (with lyrics) of 10 genres from the GTZAN dataset [17]; the genres include classical, jazz, blues, country, disco, hiphop, metal, pop, reggae, and rock.

To be consistent with the GTZAN dataset in which each music piece has a length of $30\,s$, we extract an arbitrary snippet of $30\,s$ from each of the 399 sleep music piece. All music data was preprocessed into monophonic Waveform Audio File Format (WAV) with a sampling rate of $16\,KHz$ before feature extraction and analysis.

2.2. Musical Feature Extraction

For the analysis, we consider the following typical and informative audio features:

1. Articulation feature describes the staccato and legato [18]. We extract this feature by calculating the average silence ratio (*ASR*) that indicates the percentage of frames whose root mean square (RMS) energy is lower than the average RMS energy of all frames [18, 19].

2. Energy features describe the intensity of music signals. To characterize the overall intensity and its variation throughout music, we calculate the mean, variation, and standard deviation of RMS energy (*RMS mean, RMS var, RMS std*) across all frames for each music piece.

3. Spectrum: To characterize the natural human auditory perception on a logarithmic frequency scale, we use the Mel-frequency cepstral coefficients (MFCCs) that have been proven effective in many music information retrieval tasks. For each $30\,s$ music piece, we extract the first 20 MFCCs and calculate the mean coefficient value for each of them (*MFCC 1 – MFCC 20*).

4. Rhythm characterizes the temporal note placements of a music signal. We calculate the estimated overall tempo (*Tempo*) of each music piece. In addition, to characterize the rhythm strength and its variation throughout music, we calculate the onset envelope and extract its mean (*OEnv mean*), variation (*OEnv var*), and standard deviation (*OEnv std*) as the other three rhythm features.

5. Bass and treble: We characterize these two features using the spectral rolloff frequency that stands for the frequency bin such that a certain amount of energy (85% in our study) in the current frame is obtained no higher than this frequency bin [20]. We first calculate the rolloff frequency for each frame of the given music and then calculate their mean (*SRolloff mean*), variation (*SRolloff var*), and standard deviation (*SRolloff std*) as the music features.

6. Noise- or tone-like: Spectral flatness quantifies how much a music piece is noise-like as opposed to tone-like [21]. We calculate the mean (*Flatness mean*), variation (*Flatness var*), and standard deviation (*Flatness std*) across all frames for each music piece.

We calculate the above features using Librosa[2] with a sampling rate of $16\,kHz$, FFT window size of 2048, and hop length of 512. The above 34 features were normalized before further analysis. Finally, each $30\,s$ music piece is described with a 34-dimensional feature vector.

2.3. Musical Feature Analysis

To explore which features contribute most to distinguishing sleep music from other music, we conduct an analysis using the K-means clustering technique. We specify two clusters to form, aiming to separate the 399 pieces of sleep music from the other 1000 GTZAN music/songs.

To figure out the most discriminating musical features, we calculate the metric *adjusted Rand score (ARS)* which is commonly used to measure the correctness of clustering. $ARS \in [-1, 1]$ measures the similarity between the ground-truth labels and the clusters assigned by the K-means model [22]. An ARS value of 1 means perfect matching and random assignments lead to a score close to 0. Furthermore, for a more comprehensive understanding of the clustering results, we also calculate *Silhouette Coefficient (SC)* that is commonly used to measure the compactness of clusters. An SC value of 1 means highly compact and well-separated clusters and a value of 0 indicates overlapping clusters.

Table 1 summarizes the analysis results when using different subsets of musical features. We concatenate the articulation and energy features since they are all based on RMS energies. As highlighted in Table 1, overall, spectral rolloff features perform the best to distinguish sleep music from other music. More specifically, spectral rolloff features show the highest ARS value (0.761) and the second highest SC value (0.581).

Table 1: K-means clustering results measured by *adjusted Rand score (ARS)* and *Silhouette Coefficient (SC)* when using different subsets of musical features. Each musical feature has different discriminating power, among which the spectral rolloff features perform the best to distinguish sleep music from other music.

	ARS	SC
All 34 musical features	0.115	0.424
Only articulation and energy features	-0.063	0.497
Only MFCC features	0.096	0.449
Only rhythm features	0.112	0.456
Only spectral rolloff features	**0.761**	0.581
Only spectral flatness features	0.037	0.629

To further obtain an intuitive understanding of the musical features in our dataset, we visualize the sleep music and the GTZAN dataset in 2D spaces using the t-SNE algorithm. We experiment with different perplexity factors for the t-SNE implementations, which lead to similar results, and we show an example visualization with perplexity set as 30 in Figure 1. It shows that with the exception of articulation and energy based features in Figure 1(b), most other features lead to good clustering performance for sleep music. In particular, Figure 1(e) supports our finding from Table 1 that the spectral rolloff features provide the best cluster separability for sleep music.

[1] https://www.spotify.com/us/

[2] https://librosa.org

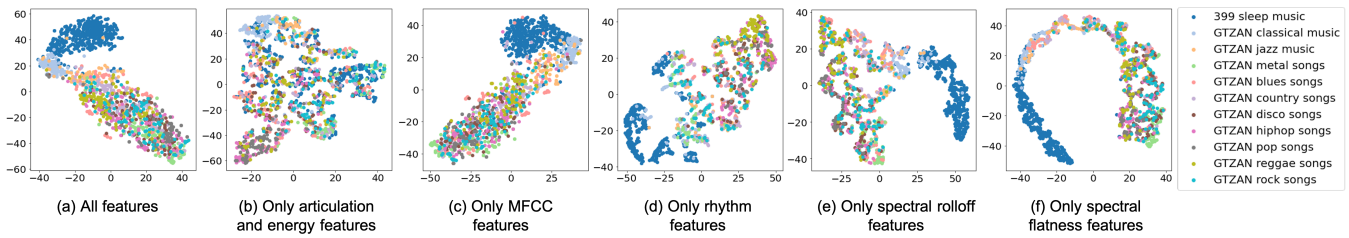

Fig. 1: Visualization of sleep music and GTZAN music when using different subsets of musical features. Compared with other musical features, the spectral rolloff features (e) group and separate the sleep music (dark blue points) more clearly from the other music.

Takeaways. Overall, the analysis in this section implies that each musical feature has a different discriminating power to distinguishing sleep music from the other music types. The results, especially the performance of the spectral rolloff features, indicate that sleep music is mostly characterized by its spectrum-relevant features that are more closely related to the bass, treble, and the overall pitch profile of a music piece.

3. SLEEPGAN: THERAPEUTIC STYLE TRANSFER

Inspired by the music style transfer technique and its application for anxiety reduction [14], we present an exploration of developing a therapeutic style transfer model called SleepGAN, which enhances the therapeutic effects of a music piece after trained using sleep music as the target style. We propose to design SleepGAN using the CycleGAN [23] structure and based on the feature analysis results in Section 2. In the following, we first present a baseline CycleGAN model, on top of which we introduce our SleepGAN that involves loss functions constructed using the 34 musical features.

3.1. Baseline: CycleGAN Model for Music Style Transfer

In our work, we regard the sleep music to be of the "therapeutic genre" that is in parallel to other music genres such as classical, pop, and electronic. Accordingly, we build our baseline style transfer model following the structure in [24] that has shown adequate performance in genre transfer. Let X denote the music dataset from the "therapeutic genre", and Y denote the music dataset from any other genre such as classical or pop. The baseline CycleGAN style transfer model consists of two generators $G : X \rightarrow Y$ and $F : Y \rightarrow X$ with a U-Net [25] architecture and two convolutional PatchGAN discriminators D_X and D_Y [26]. To train this model, the training objective

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y) + L_{GAN}(F, D_X) + \lambda_{cycle} L_{cycle}(G, F) + \lambda_{id} L_{id}(G, F) \quad (1)$$

consists of two GAN adversarial losses $L_{GAN}(G, D_Y)$ and $L_{GAN}(F, D_X)$, a cycle consistency loss $L_{cycle}(G, F)$, and an identity loss $L_{id}(G, F)$.

3.2. SleepGAN: Therapeutic Style Transfer

Section 2 reveals that different musical features contribute differently to distinguishing sleep music from the other music. Inspired by this finding, we propose to include a novel musical loss $L_{musical}(G, F)$ in SleepGAN that characterizes each of the 34 musical features according to their contribution to therapeutic effects. Formally:

$$L_{musical}(G, F) = L_{cosine}(\boldsymbol{w} \cdot f_{musical}(G(x)), \boldsymbol{w} \cdot f_{musical}(y)) + L_{cosine}(\boldsymbol{w} \cdot f_{musical}(F(y)), \boldsymbol{w} \cdot f_{musical}(x)) \quad (2)$$

in which, L_{cosine} refers to cosine similarity loss between two feature vectors, $f_{musical}$ refers to the 34-dimensional musical feature vector extracted from the current music signal, $\boldsymbol{w} = [w_1, w_2, ..., w_{33}, w_{34}]$ refers to the weight vector that weighs the 34 musical features. The weight of each feature is set to its Adjusted Rand Score (ARS) value from the K-means analysis. Thereafter, the weight vector is normalized using min-max normalization.

Including this weighted musical loss, the training objective for SleepGAN is formulated as follows:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y) + L_{GAN}(F, D_X) + \lambda_{cycle} L_{cycle}(G, F) + \lambda_{id} L_{id}(G, F) + L_{musical}(G, F) \quad (3)$$

To apply a trained baseline or SleepGAN model for style transfer, the input music is first converted into its Mel-spectrogram, then the model generator generates a new Mel-spectrogram in the therapeutic style. This output spectrogram is finally converted back to the time domain using a gradient-based inversion algorithm [27] for playback.

4. EXPERIMENTS

To evaluate the performance of our SleepGAN model, we trained a baseline and a SleepGAN model for each GTZAN genre using the 399 sleep music as the target style and a random selection of 50 music pieces from that genre as the input style. The models were trained using Adam optimizer, a learning rate of 2×10^{-4}, and batch size of 16. For the training objective, we chose parameters $\lambda_{cycle} = 1$ and $\lambda_{id} = 6$. We then tested the trained models using the rest 50 music pieces from the corresponding GTZAN genre.

4.1. Objective Performance Evaluation

One typical objective evaluation for a style transfer model is to compare the style-transferred music with respect to the reference music (i.e., sleep music in our study) by calculating their similarity using extracted features [28, 29]. In our work, we adopt the music similarity measure [30] used for music class recognition which fits the context of our study. More specifically, we extract VGGish feature vectors [31] from the sleep music and the model-generated music and use their cosine similarity as a quantitative measure. Note that we choose not to include the 34 musical features (presented in §2.2)

Table 2: Style-specific models: Mean cosine similarities with standard deviations between the sleep music and the test music of each GTZAN genre. The output music from our SleepGAN model becomes more similar to the sleep music compared to the output from the baseline model and the original version of the music.

	Original	Baseline	SleepGAN
Jazz	0.397 ± 0.069	0.476 ± 0.091	**0.529** ± 0.088
Classical	0.457 ± 0.083	0.558 ± 0.101	**0.596** ± 0.114
Rock	0.357 ± 0.083	0.430 ± 0.075	**0.451** ± 0.094
Reggae	0.368 ± 0.065	0.437 ± 0.089	**0.465** ± 0.080
Pop	0.363 ± 0.064	0.454 ± 0.103	**0.509** ± 0.067
Metal	0.323 ± 0.054	0.455 ± 0.067	**0.464** ± 0.054
Hiphop	0.352 ± 0.058	**0.444** ± 0.112	0.434 ± 0.056
Disco	0.342 ± 0.061	0.447 ± 0.065	**0.483** ± 0.076
Country	0.406 ± 0.065	**0.472** ± 0.018	**0.472** ± 0.037
Blues	0.391 ± 0.054	0.485 ± 0.068	**0.551** ± 0.049
Average	0.3756	0.4658	**0.4954**

for the similarity calculation, in order to avoid potential bias since these features are used for training the SleepGAN model.

Table 2 shows the cosine similarities between the sleep music and the test music of each GTZAN genre. The experimental results show that SleepGAN manages to change the style of the original music so as to match the style of sleep music. More specifically, overall, SleepGAN increases the cosine similarity by approximately 32% compared to the original music. We also observe that the transferred music from SleepGAN shows higher similarities than that from the baseline CycleGAN model with an increase of around 6% on average. In particular, SleepGAN successfully enhances the therapeutic style of eight music genres by approximately 8% on average. This indicates that our proposed loss function, $L_{musical}(G, F)$, well captures the therapeutic characteristics of sleep music, thereby enabling a more effective therapeutic style transfer.

We further analyze the effect of the therapeutic style transfer for different genres. Out of the 10 music genres, SleepGAN shows the higher similarity than the baseline for all genres except for hiphop and country. This is surprising, considering that music of eight genres (other than jazz and classical music) is songs with lyrics. We observe that our SleepGAN model manages to soften human voice and strong beats in the original music. According to clinical studies [10, 32], such changes could help to induce more therapeutic properties into a music piece.

4.2. Discussion

Apart from the above evaluations, we also conducted further objective and subjective assessments, as discussed below.

Universal therapeutic style transfer: So far, we have trained models for each specific input style. While this is a common approach for style transfer, such a genre-specific method brings several limitations such as limited generalizability to unseen input styles and training cost. To explore the potential of universal style transfer, we trained a baseline and a SleepGAN model using the training music from all GTZAN genres as the input style, and tested the trained models on the same test data as before. As shown in Table 3, compared with the original music that on average has a similarity value of 0.3756, the transferred music from the universal baseline model has an average similarity values of 0.4207, and from the universal SleepGAN model 0.4288. While the improvement of universal models is lower compared to that of style-specific models (see Table 2), the universal SleepGAN model still shows reasonable enhancement regardless of the genre. We argue that the model generalizability could be further advanced by integrating scalable architecture such as that of StarGAN [33].

Table 3: Style-independent universal models: Mean cosine similarities with standard deviations between the sleep music and the test music of each GTZAN genre.

	Original	Baseline	SleepGAN
Jazz	0.397 ± 0.069	0.459 ± 0.068	**0.470** ± 0.105
Classical	0.457 ± 0.083	0.477 ± 0.087	**0.542** ± 0.114
Rock	0.357 ± 0.083	0.402 ± 0.069	**0.411** ± 0.059
Reggae	0.368 ± 0.065	0.396 ± 0.077	**0.415** ± 0.081
Pop	0.363 ± 0.064	**0.424** ± 0.080	0.413 ± 0.085
Metal	0.323 ± 0.054	**0.389** ± 0.083	0.373 ± 0.089
Hiphop	0.352 ± 0.058	0.382 ± 0.056	**0.397** ± 0.056
Disco	0.342 ± 0.061	**0.375** ± 0.083	0.354 ± 0.082
Country	0.406 ± 0.065	**0.456** ± 0.067	0.451 ± 0.079
Blues	0.391 ± 0.054	0.447 ± 0.078	**0.462** ± 0.081
Average	0.3756	0.4207	**0.4288**

Subjective evaluation: To explore how users would perceive the transferred music from the models, we conducted a small-scale preliminary subjective experiment with 11 participants (9 males and 2 females). To help the participants understand what sleep music is, we first asked them to listen to three pieces of sleep music from our dataset. Then, we played 4 pieces of style-transferred music (2 randomly chosen from jazz and 2 from pop) from SleepGAN and the baseline, i.e., 8 music pieces in total, and asked the participants to rate their subjective perception of the style similarity to sleep music on a 5-point scale from 1 (not similar at all) to 5 (very similar).

We compare the scores of SleepGAN and the baseline and count which model has the higher score for the same music piece. For jazz music, SleepGAN was marked with a higher score for 40.9% of the cases, whereas the baseline was so for 22.7% of the case; for the rest of the cases, they were reported with the same score. It indicates that the participants perceived the style of the SleepGAN output more similar to the style of sleep music, compared to the output of the baseline. However, for pop music, the participants marked the higher score for the baseline (27.3%) more often than for SleepGAN (18.2%). We conjecture that this might be due to the presence of vocal elements in pop music — although our quantitative analysis showed that SleepGAN increased feature similarity post-transfer for pop music, our study participants might have focused on the vocal elements of the music and would not have noticed the change in the overall music structure. However, we are aware that the subjective evaluation was in an early phase. To better understand users' experience with the generated music, we will conduct more comprehensive studies to evaluate its musicality and clinical effects.

5. CONCLUSIONS AND FUTURE WORK

With the vision of developing personalized sleep therapy music, this paper first presented a feature analysis to uncover the impact of various musical features on the therapeutic properties of sleep music. Building on this analysis, we presented the SleepGAN model and demonstrated how to exploit the style transfer technique to automatically incorporate therapeutic properties in a user's preferred music, supported by quantitative and qualitative evaluation results.

As a future work, we are designing a large-scale clinical study to validate the efficacy of the transferred music in improving long-term sleep quality. Moreover, we are interested in enhancing the generalizability of the SleepGAN model so to be easily extended to unseen music styles.

6. REFERENCES

[1] F. P. Cappuccio, F. M. Taggart, N.-B. Kandala, A. Currie, E. Peile, S. Stranges, and M. A. Miller, "Meta-analysis of short sleep duration and obesity in children and adults," *Sleep*, vol. 31, no. 5, pp. 619–626, 2008.

[2] F. P. Cappuccio, L. D'Elia, P. Strazzullo, and M. A. Miller, "Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis," *Diabetes care*, vol. 33, no. 2, pp. 414–420, 2010.

[3] T. Trahan, S. J. Durrant, D. Müllensiefen, and V. J. Williamson, "The music that helps people sleep and the reasons they believe it works: A mixed methods analysis of online survey reports," *PloS one*, vol. 13, no. 11, pp. e0206531, 2018.

[4] "CDC declares sleep disorders a public health epidemic," https://www.sleepdr.com/the-sleep-blog/cdc-declares-sleep-disorders-a-public-health-epidemic/.

[5] L. Degenhardt, S. Darke, and P. Dillon, "GHB use among australians: characteristics, use patterns and associated harm," *Drug and alcohol dependence*, vol. 67, no. 1, pp. 89–94, 2002.

[6] D. Fancourt, A. Ockelford, and A. Belai, "The psychoneuroimmunological effects of music: A systematic review and a new model," *Brain, behavior, and immunity*, vol. 36, 2014.

[7] A. Linnemann, B. Ditzen, J. Strahler, J. M. Doerr, and U. M. Nater, "Music listening as a means of stress reduction in daily life," *Psychoneuroendocrinology*, vol. 60, pp. 82–90, 2015.

[8] U. Nilsson, "Soothing music can increase oxytocin levels during bed rest after open-heart surgery: a randomised control trial," *Journal of clinical nursing*, vol. 18, no. 15, 2009.

[9] "Bose sleepbuds," https://tinyurl.com/c778kuxs.

[10] E.-T. Chang, H.-L. Lai, P.-W. Chen, Y.-M. Hsieh, and L.-H. Lee, "The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial," *International journal of nursing studies*, vol. 49, no. 8, 2012.

[11] A. Yamasato, M. Kondo, S. Hoshino, J. Kikuchi, M. Ikeuchi, K. Yamazaki, S. Okino, and K. Yamamoto, "How prescribed music and preferred music influence sleep quality in university students," *The Tokai Journal of Experimental and Clinical Medicine*, vol. 45, no. 4, 2020.

[12] "The science behind killing a song when you listen to it too much.," https://www.independent.co.uk/life-style/killing-song-science-magic-lost-listen-too-much-sound-good-michael-bonshor-a7728156.html.

[13] G. T. Dickson and E. Schubert, "Musical features that aid sleep," *Musicae Scientiae*, p. 1029864920972161, 2020.

[14] Z. Hu, Y. Liu, G. Chen, S.-h. Zhong, and A. Zhang, "Make your favorite music curative: Music style transfer for anxiety reduction," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1189–1197.

[15] A. Yamasato, M. Kondo, S. Hoshino, J. Kikuchi, S. Okino, and K. Yamamoto, "Characteristics of music to improve the quality of sleep," *Music and Medicine*, vol. 11, no. 3, pp. 195–202, 2019.

[16] R. J. Scarratt, O. A. Heggli, P. Vuust, and K. V. Jespersen, "The music that people use to sleep: universal and subgroup characteristics," 2021.

[17] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.

[18] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation and optimization.," in *ISMIR*. Citeseer, 2008, pp. 521–526.

[19] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in *ACM SIGIR*, 2003, pp. 375–376.

[20] M. Kos, Z. Kačič, and D. Vlaj, "Acoustic classification and segmentation using modified spectral roll-off and variance-based features," *Digital Signal Processing*, vol. 23, no. 2, pp. 659–674, 2013.

[21] S. Dubnov, "Generalization of spectral flatness measure for non-gaussian linear processes," *IEEE Signal Processing Letters*, vol. 11, no. 8, pp. 698–701, 2004.

[22] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, 1985.

[23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[24] M. Pasini, "MelGAN-VC: Voice conversion and audio style transfer on arbitrarily long samples using spectrograms," *arXiv:1910.03713*, 2019.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015.

[26] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE CVPR*, 2016.

[27] R. Decorsière, P. L. Søndergaard, E. N. MacDonald, and T. Dau, "Inversion of auditory spectrograms, traditional spectrograms, and other envelope representations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 46–56, 2014.

[28] M. Tomczak, C. Southall, and J. Hockman, "Audio style transfer with rhythmic constraints," in *Digital Audio Effects (DAFx)*, 2018, pp. 45–50.

[29] O. Cífka, U. Şimşekli, and G. Richard, "Groove2Groove: one-shot music style transfer with supervision from synthetic data," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2638–2650, 2020.

[30] F. Pishdadian, P. Seetharaman, B. Kim, and B. Pardo, "Classifying non-speech vocals: Deep vs signal processing representations," 2019.

[31] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, et al., "CNN architectures for large-scale audio classification," in *ICASSP*. IEEE, 2017, pp. 131–135.

[32] H.-L. Lai, Y.-M. Li, and L.-H. Lee, "Effects of music intervention with nursing presence and recorded music on psychophysiological indices of cancer patient caregivers," *Journal of clinical nursing*, vol. 21, no. 5-6, pp. 745–756, 2012.

[33] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *IEEE CVPR*, 2018, pp. 8789–8797.