



Effect of Spectral Complexity Reduction and Number of Instruments on Musical Enjoyment with Cochlear Implants

Avamarie Brueggeman, John H.L. Hansen

Center for Robust Speech Systems, The University of Texas at Dallas, Richardson, TX, USA

avamarie.brueggeman@utdallas.edu, john.hansen@utdallas.edu

Abstract

Although speech recognition technology for cochlear implants has continued to improve, music accessibility remains a challenge. Previous studies have shown that cochlear implant users may prefer listening to music that has been reengineered to be less complex. In this paper, we consider the combined effect of spectral complexity reduction and number of instruments playing on musical enjoyment with cochlear implants. Nine normal hearing listeners rated 200 10-second music samples on three enjoyment modalities (musicality, pleasantness, and naturalness) with and without the use of cochlear implant simulation. The music samples included 20 versions of the song “Twinkle Twinkle Little Star” synthesized using one of five different instruments and with one to four instruments playing at once. The remaining 180 versions were created by reducing each sample’s spectral complexity to nine different levels using principal component analysis. The results showed a preference for less amounts of spectral complexity reduction for samples without cochlear implant simulation ($P < .001$), as well as a preference for fewer instruments for samples with cochlear implant simulation ($P < .001$). However, spectral complexity reduction was not a significant factor for samples with cochlear implant simulation, and a significant interaction effect between spectral complexity reduction and number of instruments was not found.

Index Terms: cochlear implants, music accessibility, spectral complexity reduction

1. Introduction

Hearing loss is a common disorder which affects over 460 million people worldwide [1]. For people with severe to profound levels of hearing loss, cochlear implants (CIs) may be used to help restore a sense of hearing [2]. Cochlear implants are medical devices which directly stimulate the auditory nerve via a surgically implanted array of 12-24 electrodes [3]. Electrodes are positioned tonotopically inside the cochlea in an attempt to stimulate particular frequency-dependent locations. Frames of audio from external microphones (typically worn on the CI user’s ear) are analyzed to determine the electrode activation pattern via a specific stimulation strategy [4]. In general, stimulation strategies consist of processing the input audio into a number of frequency channels and mapping them to particular electrodes to activate. Thus, the spectral resolution afforded to CI users is severely limited by the number of electrodes as well as the spectral smearing that occurs between channels [5].

Perceiving and enjoying music with cochlear implants has remained a challenge, as CI users have been shown to perform significantly more poorly on music tasks such as melody and instrument recognition compared to normal hearing (NH) listeners [6]. This is likely due to the insufficient levels of information pertinent to music that are transmitted by the implant

such as temporal fine structure, dynamic range resolution, and spectral resolution [7]. It has been shown that the number of channels necessary for speech recognition in quiet for NH listeners using CI simulation may be only 3-4 channels, although more complex listening tasks can require over 30 channels [8]. The experiment in [9] comparing the performance of NH listeners and CI users found that NH listeners required up to 32 channels for a melody recognition task. Therefore, although cochlear implants can enable a high level of speech perception (many CI users are able to achieve 90% speech recognition after two years [10]), they do not provide a similarly sufficient amount of support for music perception. Due to these issues, CI users have reported low rates of music enjoyment. In [11], CI users were asked to rate their enjoyment of music before becoming deaf and after receiving their implant; the mean scores were 8.7/10 for the former condition and 2.6/10 for the latter. Another study found that the amount of time spent listening to music and the level of music enjoyment significantly decreased after implantation, but slightly more than half of the patients still enjoyed music with their implant [12].

Recent advancements in music accessibility include strategies focused on reducing the general complexity of music to make it more enjoyable for CI users [6]. In [13], the authors created 20 less complex versions of a country music song by retaining subsets of the original instruments and had CI users and NH listeners rate them using three scales: pleasantness, musicality, and naturalness. Overall, CI users and NH listeners with CI simulation found reduced versions with 1-3 instruments more enjoyable than the original version. Another study found that CI users preferred music with one instrument over multiple instruments as well as music with a relatively simple structure over a complex structure [7]. Reducing the harmonic series was used in [14] to create four versions of the song “Happy Birthday” with decreased spectral complexity. In this study, the authors synthesized the song using seven different instruments and found that NH listeners with CI simulation and CI users preferred the most harmonically reduced versions, rating them as more pleasant and/or natural.

A number of papers investigated the use of dimensionality reduction techniques to reduce spectral complexity. In [15, 16, 17], the authors computed reduced rank approximations via principal component analysis (PCA) and partial least-squares analysis of music spectra using the constant-Q transform (CQT). They evaluated their strategies with CI users in [16] and found that the spectrally reduced versions were significantly preferred. In [17], the authors created an interactive interface which allowed CI users to select their preferred level of spectral complexity (via number of retained principal components) while listening to music in real time. The spectral complexity reduction was performed using the overlap-add method with PCA on a sliding window CQT. A variety of 10-second classical music samples were used which contained voices for

the melody and accompaniment, and the results indicated a significant preference for samples with high levels of spectral complexity reduction. The mean number of preferred principal components to retain was between 3.5 and 7.5 for the 16 classical music samples.

Considering these results, we were interested in exploring whether a relationship existed between spectral complexity reduction and number of instruments on music enjoyment; in particular, we hypothesized that a music piece with fewer instruments may require less spectral complexity reduction to be enjoyable for CI users due to it already being less complex. To this end, we designed a listening experiment described as follows.

2. Materials and Methods

2.1. Participants

Nine (N=9) normal hearing (NH) listeners took part in the music listening experiment. The participants were between the ages of 23-30 years (mean: 26.7 years, standard deviation: 2.5 years). Three participants were female and six were male. No participant had a history of hearing loss. Four participants had received at least one year of formal music training (range: 1-10 years, mean: 4.3 years, standard deviation: 3.4 years). All participants were fluent in English.

2.2. Cochlear implant simulation

To simulate the sound of a cochlear implant, a 22-channel pulse-spreading harmonic complexes (PSHC) vocoder [18, 19] was used to approximate the spectral smearing and reduced spectral resolution experienced by CI users. The decision to use a PSHC vocoder was based on the results of a study comparing noise, sine, and PSHC vocoders which found that CI users with single-sided deafness rated the PSHC vocoder as sounding the most similar to a CI [20]. Our vocoder logarithmically spaced its 22 channels from 188 to 7,938 Hz based on the Nucleus CI24 cochlear implant system (Cochlear Limited, Sydney, Australia) which uses 22 electrodes. The vocoder also selected a subset of channels per frame for synthesis based on the advanced combinatorial encoder (ACE) strategy [21] used in Nucleus implants. During ACE, an n -of- m method is employed which selects a number of channels (typically 8-10) out of 22 for stimulation per frame based on channels with the largest envelope amplitude. Our implementation used $n = 8$.

2.3. Listening experiment

The listening experiment was implemented using a full factorial design with two factors: spectral complexity reduction (10 levels) and number of instruments playing (4 levels). The goal of the experiment was to examine the possible main effects and interaction effects of these factors on musical enjoyment. For music samples using CI simulation, our initial hypothesis was that a potential interaction effect might exist where music that contained fewer instruments would not require as much spectral complexity reduction as music with more instruments to be perceived as enjoyable due to these samples already being less complex.

To this end, a music dataset was created which consisted of 200 10-second music samples of “Twinkle Twinkle Little Star.” Each sample was created with a particular level of spectral complexity and number of instruments. The samples were synthesized using five different instruments selected from various musical families such as woodwind and brass. The song “Twinkle

Twinkle Little Star” was chosen due to its relatively simple structure and because its melody would already be familiar to most participants.

During the experiment, participants rated each music sample on three visual analog scales from 0 to 10 to assess musical enjoyment: musicality (“sounds like music” to “does not sound like music”), pleasantness (“sounds pleasant” to “does not sound pleasant”), and naturalness (“sounds natural” to “does not sound natural”). In each case, a higher score indicated a more favorable rating. Each music sample was played on loop while being rated. Participants could pause and resume listening to the sample any number of times. Participants could not go back to a previous music sample once it had been evaluated.

Music samples with the same number of instruments playing were presented together in blocks. The order of the blocks and the order of samples within the blocks were randomized within and between subjects. The experiment was immediately repeated using CI simulated (vocoded) samples with a similar randomization strategy. In every case, participants experienced the non-vocoded part of the experiment first and the vocoded part last. There was not a practice session. In total, participants listened to 400 unique music samples (200 non-vocoded, 200 vocoded). Participants completed this experiment as a take-home task during which they were asked to wear headphones and listen to the samples in a quiet environment. The graphical user interface for the experiment was created using App Designer in MATLAB 2019b (The MathWorks, Inc, Natick, Massachusetts).

2.4. Music dataset

The music dataset was created based on the traditional nursery rhyme “Twinkle Twinkle Little Star.” A four-part MIDI score, shown in Figure 1 for piano, was used to generate four versions of the song with 1-4 of the topmost instruments. For example, in Figure 1, the 1-instrument version would contain Piano 1, the 2-instrument version would contain Piano 1 and 2, and so on. Note that the first instrumental part carries the main melody and contains the same rhythm as the second instrumental part (the third and fourth instrumental parts also share the same rhythm). Therefore, each of the four versions of the song contained the main melody with zero to three harmonizing voices.

Each song version was synthesized with five different instruments in MuseScore 3 (MuseScore, Sint-Denijs-Westrem, Belgium) using the default instrument library. The selected instruments were chosen to represent a diverse range of musical sound from different musical families: B-flat clarinet (woodwind), B-flat trumpet (brass), piano (keyboard), violin (bowed string), and marimba (pitched percussion). All instruments played the song in A-flat major (sounding pitch) at 100 beats per minute; therefore, each instrumental part was played at the same frequency across all instruments. To keep the sample length brief, only the first four measures of “Twinkle Twinkle Little Star” were selected and synthesized as 44.1 kHz stereo WAV files for each combination of song version and instrument, resulting in 20 10-second samples. Each sample was converted from stereo to mono and resampled to 16 kHz.

2.5. Spectral complexity reduction

To create 10 levels of spectral complexity reduction, we followed the general method described in [17] which used PCA to create reduced rank approximations of CQT spectra. As in [17], a first order pre-emphasis filter was applied to all samples to counteract the low-pass filter effect which can occur in audio



Figure 1: First four measures comprising the 10-second excerpt of “Twinkle Twinkle Little Star” with four piano parts.

samples after retaining a small number of principal components using PCA.

2.5.1. Constant-Q transform

The spectral representation of each sample was computed using the CQT [22]. Unlike the Fourier transform, the CQT uses a variable frame size across frequency bands which enables it to attain higher spectral resolution at lower frequencies and higher temporal resolution at higher frequencies [23, 24], mimicking human hearing. It uses logarithmically spaced frequency bins which maintain a constant ratio of center frequency to bandwidth. We applied the CQT using 24 frequency bands per octave from 55 to 7,040 Hz as in [17]. This resulted in 168 frequency bands over the 7 octaves.

For each sample’s CQT spectrum, dimensionality reduction was applied via PCA using a sliding window of 240 samples (approximately 600 ms) with 50% overlap. The window size was chosen based on the length of the shortest note (quarter note) found in the music piece (600 ms). In [15], the authors chose the window size based on the median note length from numerous pieces of music; however, since our song “Twinkle Twinkle Little Star” only contains quarter notes and half notes, we simply used the length of the former note.

2.5.2. Principal component analysis

To reduce the spectral complexity of each sample, PCA was used to reduce the dimensionality of the CQT spectrum by retaining a limited number of principal components [25]. The retained components were then used to create a reduced rank approximation of the original spectrum. Ten levels of spectral complexity were generated by either performing no reduction (original sample) or retaining the first 20, 10, 8, 7, 6, 5, 4, 3, or 2 principal components. The inverse CQT was used to revert back to the time domain. Finally, a de-emphasis filter was used to complete the emphasis process.

In general, samples created by retaining fewer principal components were considered to be less spectrally complex because the approximated CQT spectrum would contain less low variance content. For a music signal with a melody and accompaniment, it is assumed that the melodic part would generally

contain the most prominent harmonics [15]. Therefore, PCA will preserve dominant parts of the music (mainly the melody) while suppressing minor parts of both the melody and accompaniment. We chose the number of principal components to retain based on those used in [17] but replaced 15 with 2 components because of the strong preference for a small number of retained components.

3. Results

A linear mixed effects model with random intercepts was used to analyze the effect of spectral complexity reduction and number of instruments on each of the three musical enjoyment scores for the case of both regular and CI simulated music samples. The spectral complexity condition, number of instruments condition, and their interaction term were treated as fixed effects, while subject and instrument were treated as random effects. A simulated likelihood ratio test with 1,000 replications was used to test for significant fixed effects. All calculations were run using the Statistics and Machine Learning Toolbox in MATLAB 2019b.

The mean scores on each enjoyment scale for samples with and without CI simulation are shown in Figure 2 and 3 as a function of retained principal components and number of instruments, respectively. Samples without CI simulation attained higher mean scores than samples with simulation for all three enjoyment measures. A significant interaction effect between spectral complexity and number of instruments was not found for any enjoyment scale.

3.1. Normal hearing listeners without CI simulation

The spectral complexity condition was found to affect all three scores ($P < .001$) for samples without CI simulation such that an increase in spectral complexity reduction was associated with a decrease in score. In particular, moving from samples with no spectral reduction to samples with maximal reduction (retaining 2 principal components) led to a decrease in score of approximately 1.03 ± 0.18 (standard error) for musicality, 1.68 ± 0.20 for pleasantness, and 2.29 ± 0.20 for naturalness. In addition, the number of instruments was found to be a significant fixed effect for musicality ($P = .002$) and pleasantness ($P < .001$).

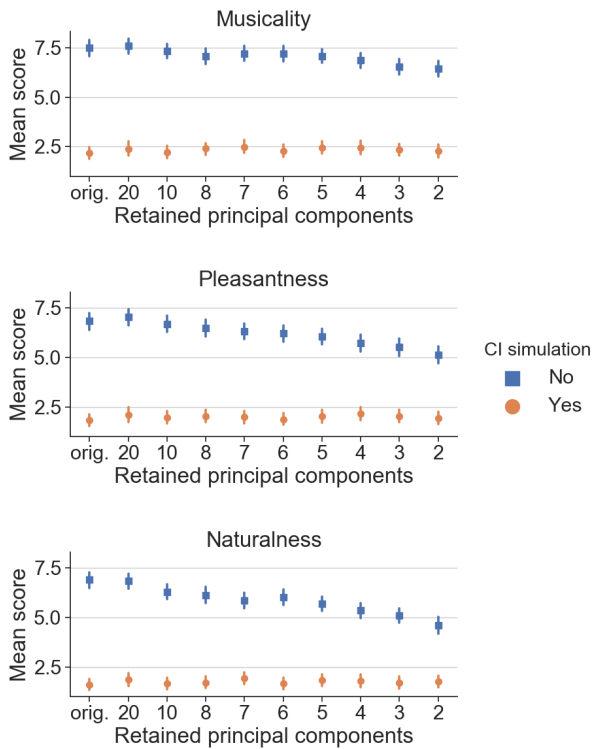


Figure 2: Mean scores with 95% confidence intervals for musicality (top), pleasantness (middle), and naturalness (bottom) as a function of retained principal components across all instruments and number of instruments.

3.2. Normal hearing listeners with CI simulation

For all three scores, the number of instruments was shown to be a significant fixed effect ($P < .001$) for samples with CI simulation. Increasing the number of instruments from 1 to 4 was associated with a decrease in score of approximately 1.21 ± 0.10 (standard error) for musicality, 0.93 ± 0.10 for pleasantness, and 1.16 ± 0.10 for naturalness. Spectral complexity was not found to be a significant fixed effect for any of the enjoyment scales when using CI simulation.

4. Discussion

The preference of NH listeners with CI simulation for music samples with fewer numbers of instruments agrees with findings in previous studies [7, 13]. However, unlike [15, 16, 17], no significant relationship between score and spectral complexity reduction with CI simulation was found. This may be due to factors such as the complexity of the song chosen for this experiment. Whereas the music dataset in the aforementioned works contained classical music, the relatively simple spectral structure of “Twinkle Twinkle Little Star” may have produced negligible differences between spectral complexity levels, resulting in similar ratings. In addition, [16, 17] obtained data using CI users, whereas we used NH listeners with CI simulation. More research may be required to consider the effectiveness of spectral complexity reduction for different genres of music on CI users and NH listeners with CI simulation.

Our results also found that NH listeners without CI simula-

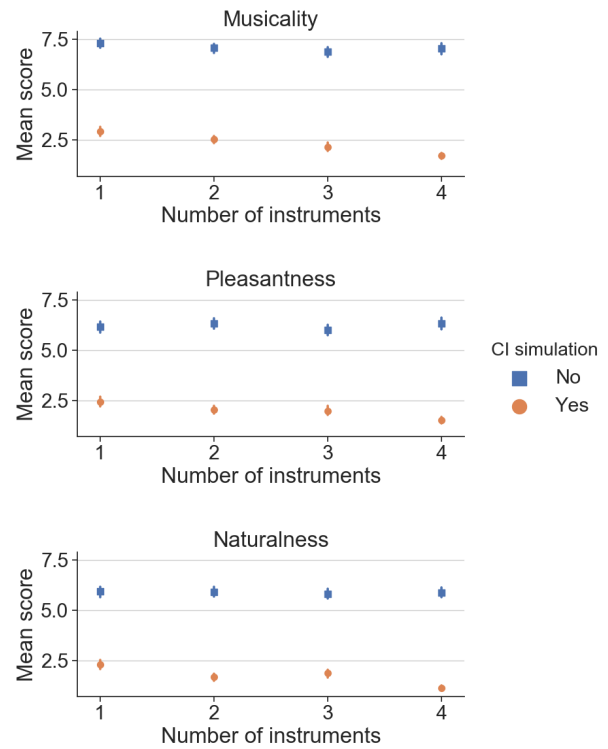


Figure 3: Mean scores with 95% confidence intervals for musicality (top), pleasantness (middle), and naturalness (bottom) as a function of number of instruments across all instruments and retained principal components.

tion preferred samples with less spectral complexity reduction. In particular, increasing the level of spectral complexity reduction had the largest negative effect on the perceived naturalness of a music sample. This result was expected considering that NH listeners are generally able to perceive more acoustic information than CI users and therefore do not require complexity reduction for music.

Future research will consider more genres of music as well as extend the number of instruments condition to include music with and without voice. We will also collect subjective ratings from both NH listeners and CI users and compare their results.

5. Conclusion

NH listeners using CI simulation were found to prefer music with fewer instruments, while less spectral complexity reduction was preferred for NH listeners without CI simulation. However, NH listeners using CI simulation did not show a significant preference for music with its spectral complexity reduced. In addition, we did not find a significant interaction effect between spectral complexity reduction and number of instruments for any enjoyment measure. More research may be needed for the effect of spectral complexity reduction on simple pieces of music.

6. Acknowledgements

This work was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1746053.

7. References

- [1] "Deafness and hearing loss," Mar 2020. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- [2] "Cochlear implants," Mar 2017. [Online]. Available: <https://www.nidcd.nih.gov/health/cochlear-implants>
- [3] A. Dhanasingh and C. Jolly, "An overview of cochlear implant electrode array designs," *Hearing Research*, vol. 356, pp. 93 – 103, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378595517302940>
- [4] C. Choi and Y.-H. Lee, *A Review of Stimulating Strategies for Cochlear Implants*, 04 2012.
- [5] D. Landsberger and A. Srinivasan, "Virtual channel discrimination is improved by current focusing in cochlear implant recipients," *Hearing research*, vol. 254, pp. 34–41, 05 2009.
- [6] W. Nogueira, A. Nagathil, and R. Martin, "Making music more accessible for cochlear implant listeners: Recent developments," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 115–127, 2019.
- [7] C. J. Limb and A. T. Roy, "Technological, biological, and acoustical constraints to music perception in cochlear implant users," *Hearing Research*, vol. 308, pp. 13 – 26, 2014, music: A window into the hearing brain. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378595513001020>
- [8] R. Shannon, Q.-J. Fu, and J. Galvin, "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta oto-laryngologica. Supplementum*, vol. 124, pp. 50–4, 06 2004.
- [9] Y.-Y. Kong, R. Cruz, J. Jones, and F.-G. Zeng, "Music perception with temporal cues in acoustic and electric hearing," *Ear and hearing*, vol. 25, pp. 173–85, 05 2004.
- [10] B. Wilson and M. Dorman, "Cochlear implants: A remarkable past and a brilliant future," *Hearing research*, vol. 242, pp. 3–21, 07 2008.
- [11] S. Mirza, S. Douglas, P. Lindsey, T. Hildreth, and M. Hawthorne, "Appreciation of music in adult patients with cochlear implants: A patient questionnaire," *Cochlear implants international*, vol. 4, pp. 85–95, 06 2003.
- [12] L. Lassaletta, A. Castro, M. Bastarrica, R. Pérez-Mora, R. Madero, J. De Sarriá, and J. Gavilán, "Does music perception have an impact on quality of life following cochlear implantation?" *Acta oto-laryngologica*, vol. 127, pp. 682–6, 07 2007.
- [13] G. Kohlberg, D. Mancuso, D. Chari, and A. Lalwani, "Music engineering as a novel strategy for enhancing music enjoyment in the cochlear implant recipient," *Behavioural Neurology*, vol. 2015, pp. 1–7, 10 2015.
- [14] J. Nemer, G. Kohlberg, D. Mancuso, B. Griffin, M. Certo, S. Chen, M. Chun, J. Spitzer, and A. Lalwani, "Reduction of the harmonic series influences musical enjoyment with cochlear implants," *Otology & Neurotology*, vol. 38, p. 1, 10 2016.
- [15] A. Nagathil, C. Weihs, and R. Martin, "Spectral complexity reduction of music signals for mitigating effects of cochlear hearing loss," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 445–458, 2016.
- [16] A. Nagathil, C. Weihs, K. Neumann, and R. Martin, "Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 142, no. 3, pp. 1219–1228, 2017. [Online]. Available: <https://doi.org/10.1121/1.5000484>
- [17] J. Gauer, A. Nagathil, R. Martin, J. P. Thomas, and C. Völter, "Interactive evaluation of a music preprocessing scheme for cochlear implants based on spectral complexity reduction," *Frontiers in Neuroscience*, vol. 13, p. 1206, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2019.01206>
- [18] E. Gaudrain and D. Başkent, "Factors limiting vocal-tract length discrimination in cochlear implant simulations," *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1298–1308, 2015. [Online]. Available: <https://doi.org/10.1121/1.4908235>
- [19] G. Hilkuysen and O. Macherey, "Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering," *The Journal of the Acoustical Society of America*, vol. 136, no. 3, pp. 1281–1294, 2014. [Online]. Available: <https://doi.org/10.1121/1.4890642>
- [20] C. Karoui, C. James, P. Barone, D. Bakhos, M. Marx, and O. Macherey, "Searching for the sound of a cochlear implant: Evaluation of different vocoder parameters by cochlear implant users with single-sided deafness," *Trends in Hearing*, vol. 23, p. 233121651986602, 09 2019.
- [21] J. J. Kiefer, S. Hohl, E. Stürzebecher, T. Pfennigdorff, and W. Gstöttner, "Comparison of speech recognition with different speech coding strategies (speak, cis, and ace) and their relationship to telemetric measures of compound action potentials in the nucleus ci 24m cochlear implant system," *Audiology : official organ of the International Society of Audiology*, vol. 40 1, pp. 32–42, 2001.
- [22] J. C. Brown, "Calculation of a constant q spectral transform," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991. [Online]. Available: <https://doi.org/10.1121/1.400476>
- [23] G. A. Velasco, N. Holighaus, M. Doerfler, and T. Grill, "Constructing an invertible constant-q transform with nonstationary gabor frames," 09 2011.
- [24] N. Holighaus, M. Doerfler, G. A. Velasco, and T. Grill, "A framework for invertible, real-time constant-q transforms," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, 09 2012.
- [25] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," Nov. 1901. [Online]. Available: <https://doi.org/10.1080/14786440109462720>