Radboud University Nijmegen

Faculty of Social Sciences

A case-based music recommendation system for cochlear implant users

Thesis BSc Artificial Intelligence

Supervisor:
Yagmur Güçlütürk

Author:
Andre Tanuri Vargas

Second reader:
Eva van den Borne

May 2021

Abstract

Cochlear implant users face many difficulties in their music listening experience, such that music perception and music appreciation remain a challenge. Rhythm is the musical element that is better perceived by CI users when compared to pitch, melody, and timbre. The presence of lyrics and percussive sounds have a positive influence on the CI users' music appreciation. Music recommendation systems can give appropriate song suggestions to a user. However, the available music recommendation systems are based on timbre and melodic content. In this paper, I propose a case-based music recommendation system that considers music perception and music appreciation from a CI user perceptive.

Contents

1 Introduction

Music is one of the oldest art forms, remaining of great importance in many of today's prevailing cultures for being fundamental in religious rites, entertaining activities, and social events and ceremonies. Listening to music can become an integral aspect of a person's life, one of the reasons why people with severe to profound hearing loss might struggle to adapt to a quiet-sounding world. Part of this population resort to neuroprosthetic devices, hoping to restore their sense of hearing and their ability to listen to music. Despite the undeniable advances in such technologies, many improvements to the aided musical experience remain an open challenge. In the field of music information retrieval, intelligent systems aim to provide better experiences to their users. Music recommendation systems aim to suggest appropriate items to the user by acknowledging user needs and domain knowledge. Considering that the available music recommendation systems reflect the necessities of normal-hearing people, an adapted music recommendation system might be beneficial for the cochlear implant users' music experience.

1.1 The cochlea

Absolute silence is impossible [96]. Such phenomenon would only exist in the total absence of particles, a perfect vacuum, which does not occur naturally in our world and has not yet been reproduced in a laboratory for longer than an infinitesimal moment [59]. While the idea of an environment at total silence can be conceived at an ideal level, the reality of the physical universe is that it is filled with sounds. While clapping hands to the beat of a song, for example, a physical force emerges from the point of contact of the palms, altering the equilibrium of the atoms in the air immediately next to it. Whenever the atoms of a medium, such as water or air, have their baseline equilibrium disrupted by a sound source, a sequence of changes in atomic pressure is unleashed as a result of the initial pressure alteration [78]. The atoms closer to the sound source are compressed together as a response to the force, raising the local atomic pressure. This is followed by a period of decompression and lower atomic pressure, which in turn is proceeded by another period of compression. In that manner, by oscillating between high and low atomic pressure, vibrations propagate from the sound source and travel through the medium as an acoustic wave [4, 1].

As human beings, we have successfully developed the capacity to perceive a range of these pressure oscillations as sound [111]. In the auditory system, sound waves are captured by the outer ear and directed towards the tympanic membrane. The thin tissue vibrates as a response to the pressure oscillations in the neighboring atoms of the medium, producing a signal in the form of mechanical movement. The ossicles connect the vibrating tympanic membrane to the oval window in the bony labyrinth, transmitting the mechanic signal by disturbing the atomic pressure of the perilymph. The sound wave traverses the perilymph into the cochlea until it reaches the endolymph inside the cochlear duct. Hair cells from the Organ of Corti detect the wave motion in their environment via a set of projections touching the surrounding endolymph. The mechanical displacement of these projections stimulates the hair cells to discharge an electric signal that proceeds to the brain via the auditory nerve, ultimately leading to sound perception.

Hair cells can also detect particularities of the physical characteristic a sound wave, such as its amplitude and frequency. The amplitude of an acoustic wave is the amount of atomic pressure deviation from the average in the medium, measured in Pascal (Pa) units. Vibrations with higher amplitudes stimulate more hair cells, as well as a faster rate of electrical discharge. It is also possible to characterize the hair cells by the extent to which they respond speeds of vibration. When measured in Hz, frequency is the number of times the sound wave returns to the starting point of the oscillations in atomic pressure during a second. Each bundle of hair cells along the cochlea can detect a specific fraction of all detectable frequencies, a consequence of the varying stiffness of the resonant basilar membrane underneath the organ of Corti. The frequency information in the sound wave's spectrum is analyzed and mapped into an organized pattern of spatiotemporal discharges by the hair cells.

When measuring cochlear activity in response to a pure tone - a sound wave that oscillates at an exact single frequency - it is possible to observe a peak of electric activity in a correspondent area of the cochlea [91, 112, 46]. Hair cells located at the base of the cochlea respond to faster frequencies (more oscillations in a second), whereas hair cells at the apex respond to slower frequencies (fewer oscillations in a second). A shift in frequency is reflected as a displacement in the cochlea's area of activity. Another mechanism underlying the encoding of sound waves into electric discharges is the rate at which the hair cells spike as a response to the incoming sound [14], since the patterns in temporal discharges also provide information about the structure of the acoustic wave.

It has also been proposed that the cochlea functions as a set of auditory band-pass filters, where each filter accentuates a continuous range of frequencies while attenuating others [36]. However, two sound frequencies that are close together in the same bandwidth and presented at the same time can interfere with each other, a phenomenon known as simultaneous auditory masking [39]. For instance, a sound wave at 320 Hz with greater amplitude may hinder the perception of another sound at 330 Hz [39].

Most of the time, sounds from the environment are not pure tones. Instruments, for example, have a fundamental frequency that is dominant [53, 52, 15], but other frequencies are also present in the formation of the sound wave [43, 3]. The relative proportion of these multiple underlying frequencies uniquely shapes the acoustic signal, allowing sound sources to be identified and differentiated from each other [69]. The specialized groups of hair cells fire in different regions along the cochlea proportionally to the multiple underlying frequencies of the acoustic signal [38]. Such a capacity of the hair cells to respond to sound at a frequency level, reveals why the proper functioning of the hair cells has a central role in the perception and characterization of sounds. For that reason, the impairment of hair cells leads to degraded hearing and, depending on the gravity of the functional impairment, total inability to perceive sounds [100].

1.2 Cochlear implants

Natural regeneration of the hair cells is a lacking ability of the inner ear [21]. However, advances in science have allowed for the development of technologies that aim to restore hearing. A cochlear implant (CI) is an effective treatment for people with severe to profound hearing loss due to damage or deficiency of the hair cells [44, 17, 119, 65]. The implant is an electronic neuroprosthetic device that requires surgical intervention for the allocation of one of its components along the cochlea. The device is able to transduce acoustic signals into a sequence of electric pulses that directly stimulate the auditory nerve, resulting in users perceiving sound [113]. The CI's external component consists of a microphone and a signal processor behind the ear for capturing and analyzing sounds from the environment. The external component connects to the internal component via a coil coupled to the scalp which is responsible for transmitting the processed signal. The internal component comprises an antenna, a magnet, and an electrode array placed along the cochlea. An encoding strategy converts the signal into a pattern of electric pulses that are addressed to the electrodes. In order to accomplish this mapping, the sound processor carries the analysis of the sound by either performing a Fourier decomposition or by directly applying band-pass filters, with the intent of estimating frequency levels at different bands of the frequency spectrum. Such analysis allows for an estimation of the intensity and rate of electrical activation that should proceed to the intracochlear electrodes, where the estimated levels of each frequency band are addressed to the correspondent electrode surfacing the area of the cochlea responsible for detecting that frequency band.

Since its early versions implanted at the end of the 1970s, the CI technology has undergone a series of improvements - from reductions in the size and multiplication of the number of electrodes to more efficient sound encoding and processing techniques. These improvements have allowed CI users to obtain satisfactory levels of speech recognition, at least in favorable conditions where background noise is acceptable [98, 85, 25]. Still, the sound perceived with electrical stimulation does not resemble the one perceived with functioning hair cells. The areas of the auditory nerve stimulated by the electrodes are broad and overlapping, and the limiting number of electrodes can only convey a poor estimate of the frequency spectrum information [2, 77]. Consequently, information from complex sound waves that require fine-grained frequency processing may get lost during transduction.

1.3 Music with cochlear implants

Music compositions are complex acoustic signals, for they are an organized arrangement of multiple sounds over time. The perception of a single musical unit requires the simultaneous and consecutive processing of occurring sounds[55]. CI users report degraded quality of music conveyed by the implant, especially when compared to the sound of music perceived before hearing loss [29]. As expected, this has an enormous impact on the CI user's music listening experience and listening habits [68], such that they believe that a better music listening experience can lead to an improved quality of life [20].

Nevertheless, some users do report listening to and appreciating music in their daily life [117, 74]. To understand the complexity of listening to music with cochlear implants, it is important to investigate the experience that the CI user has when listening to music. Due to little correlation between perception levels and appreciation levels in musical tasks, both factors can be analyzed separately as being contributors to the music experience of the CI user [19, 114]. Overall, studies show that both music perception and appreciation levels are considerably lower for CI users compared to before their hearing loss and also compared to normal-hearing people.

1.3.1 Perception

It is undeniable that music compositions comprise a variety of characteristics and patterns of sounds. Still, it is possible to identify four primary elements in the great majority of songs, namely pitch, melody, timbre, and rhythm [63]. These elements emerge from a wide range of instruments, including the human voice. The arrangement complexity of these instrument sounds differs according to the music genre in which the song is composed [83]. The ability to accurately perceive music's four primary elements allows for identifying genre and distinguishing between two songs. Due to the limited sound quality conveyed by the implant, CI users face difficulties when trying to identify different music genres or recognizing instruments in a piece of music [45].

Moreover, in order to assess the degree to which CI users are able to perceive music, researchers resort to measuring the users' performance in perception tests involving the four primary elements of music. Music perception depends on the complex and simultaneous interplay of pitch, melody, timbre, and rhythm. In tasks where it is necessary to recognize an excerpt from a full song with multiple instruments, it is clear that all primary elements are involved. Still, a common experimental approach to understand the CI user's music perception is to measure the perception accuracy for each of these musical elements separately. In general, CI users have significantly lower accuracy in music perception compared to normal-hearing people for pitch, melody, and timbre. On the other hand, CI users' performance in tasks of rhythm perception is comparable to normal-hearing people.

Pitch

Pitch is the subjective perceptual property of sound ruled by the fundamental frequency at which a sound wave oscillates [79]. Music notes represent sound waves with a fundamental frequency between 20 Hz and 20 kHz in an ordered and structured scale. Faster oscillations yield higher pitches, whereas slower frequencies fall in the lower pitch range. The most commonly used scale in Western music is the chromatic scale, consisting of twelve pitch intervals, each spaced by a semitone. Most of the instruments used in Western music are fabricated and tuned to produce these twelve semitones in different octaves. For example, the universal tuning standard sets the music note A4 at 440 Hz and A5, an octave higher, at an exact double of the oscillation speed. Not only the pitch perceived by the CI user is deviant from the actual pitch of the auditory stimulus by at least an octave, but also the total frequency range perceived by the users is halved [120].

Researches have proposed a variety of tests to evaluate pitch perception of both normal-hearing people and CI users. In pitch interval discrimination tests, CI users' are presented with a pair of music notes and need to determine if the two sounds are different, regardless of the direction of the change. One study investigated the number of semitones necessary for a person to perceive a difference in pitch [110]. The CI users' average was 5.5 semitones, whereas the normal-hearing average was 0.4 semitones. Performance within the CI users' group varied widely, from 0.8 to 19.6 semitones. Another study investigated the pitch discrimination ability of fifteen CI users along with nine normal-hearing subjects [66]. CI users scored significantly lower compared to normal-hearing subjects. Implant users performed at a chance level on the discrimination of an interval of three semitones.

In pitch ranking tests, participants are presented with a pair of music notes, but this time they need to indicate which of the two notes has a higher pitch. One study compared multiple sound processing strategies for comparing the pitch ranking ability of the CI user [109]. The participants of the study listened to two sung vowels with an interval of six semitones. The subjects' performance was, on average, poorer than normal-hearing people's, even though the study points the possible influence of different strategies in pitch ranking. Another pitch ranking study tested 114 cochlear implant users and twenty-one normal-hearing participants indicated the direction of the change in pitch [33]. Baseline fundamental frequencies ranged between 131 Hz and 1048 Hz and the intervals with respect to the baseline ranged from 1 to 4 semitones. Similar to the other studies, the performance of CI users was significantly lower than the performance of normal-hearing participants.

Melody

Melodies are closely related to pitch, as they are combinations of successive music notes perceived as a continuous unity. In the formal structure of music compositions, instruments often follow musical phrases, which are recurrent melodies salient in the song. Identifying the relative direction and range of intervals is essential to processing the nuances of melodic arrangements [72]. Some melodies are easily recognizable for being broadly known in folk culture. A good example is the song "Happy Birthday" which has variations in different languages but still follows the same identifiable melody. Other songs like "The Wedding March" are recognizable by their distinguishable melody, without any lyrical content in their composition. One study measured the ability of 49 CI users and twenty normal-hearing adults to recognize twelve familiar melodies [32], with songs presented as a sequence of music notes produced by the same instrument. Results showed that normal-hearing people were significantly more accurate than CI users, surpassing their mean accuracy in recognizing familiar songs in over 70 percent.

Musical chords, a set of harmonic notes played simultaneously, are another central constituent of melodies. Being able to identify multiple music notes is central to the perception of chords. A study showed that CI users performed near to chance level when differentiating between two or three pitches played simultaneously [18]. The specific ordering of the music notes in chord progressions accounts for the sense of melodic resolution known as cadence. Cadences can elicit either pleasant or unpleasant emotional responses, depending on the harmonic expectancy created with the initial chords in the progression [102, 48]. Another study assessed the judgment of normal-hearing people and CI users in determining whether a final chord in a progression elicits a resolution. CI users showed a diminished capacity to categorize such cadences compared to normal-hearing people.

Timbre

The ability to recognize instruments by their sound is rooted in the capacity to identify different timbres. Timbre is the quality of sound that allows for the distinction of equally pitched tones produced by distinct sound sources. A piano and a flute playing an A4 are perceived as two distinct sounds. On top of the fundamental frequency of 440 Hz, multiples (e.g. 880Hz, 1320Hz) known as harmonics also form the acoustic wave. The relative strength of these harmonics to the fundamental frequency is what shapes timbre.

Different timbres are known to elicit emotional responses during the process of music perception [40, 121, 8]. Several studies reported the difficulties of CI users in the perception of timbre and the identification of distinct sound sources in music [71]. A study investigated the recognition of melodies with the same notes played by eight distinct instruments from four families. CI users' were significantly less accurate than normal-hearing people on timbre recognition for all the instrument families (47% against 91% mean score for instrument identification) [34]. As expected, the difficulty in perceiving different instruments reflects in the description of real-world excerpts from songs, in which there are multiple timbres entangled in a single sound wave [31]. Finally, another study noticed that some timbres are more easily recognizable, which is the case for the piano and drum sounds since CI users were more accurate in naming those instruments [75].

Rhythm

The perception of rhythm emerges from the capacity to extract information from patterns of sounds. The sounds' relative position in time, as well as their duration and amplitude, are central to the perception of rhythm [27]. Tempo is the speed of a music piece, constituted of an underlying temporal unit that guides all instruments in a song, often measured in beats per minute (BPM). Repetition is crucial to the perception of this song meter, as well as rhythmic regularities with contrasting sections [101]. Modulations with frequencies from 0 Hz and 10 Hz, which are below the frequency range perceived as pitch, are perceived as rhythm. For instance, when a song has a tempo of 120 BPM, this can be interpreted as $120/60 = 2$ modulations per second (2 Hz).

Researchers have proposed many tasks for evaluating the CI users' rhythm perception. Two studies presented pairs of distinct rhythmic patterns, represented as sequences of equally pitched tones, asking CI users and normal-hearing people to indicate whether the two rhythmic patterns were the same [30, 35]. These studies also tested CI users for recognition of melodic similarity. Results suggested that CI users are more accurate in perceiving similarity in rhythmic patterns than perceiving similarities in melodies. This finding contrasts with the performance of normal-hearing people, which were more accurate at estimating similarity in melodies. Instruments with sharp attacks - a quick onset in the pressure intensity of the sound wave followed by a fast decay - are the most well perceived by cochlear implants, especially when compared to soft attack instruments, such as violins, cellos and flutes [70]. CI users were also accurate in capturing the underlying meter of a song, considering they can detect when percussive sounds deviate from the underlying metric unit [51]. This finding might explain how CI users can move their bodies in synchronicity with the beat of a song [84]. These studies suggest that, from all the primary musical elements, rhythm is the most well perceived by CI users.

1.3.2 Appreciation

Another noticeable factor that plays a role in the CI users' music listening experience is music appreciation. The majority of studies focused on self-reported questionnaires to evaluate the music appreciation of CI users. Although appraisal for listening to music is lower with a CI when compared to listening to music before hearing loss, CI users still report appreciation for music that is better than neutral [76].

Another study supported that familiarity with a song has a positive influence on likability, especially in the case of popular music [28]. Regarding families of instruments, CI users reported greater appreciation for the piano, with a significant difference for the other families that were less appreciated [34]. Slow rhythm and music with lyrics may improve the music listening experience of CI users considering they were the most well-appreciated features of music [75]. This study also supported the idea that familiarity with a song is an enhancer of music enjoyment, as almost 80% of the participating CI users rated familiarity with the music and lyrics as the main factor behind their music appreciation.

1.4 Music recommendation systems

Music is one of the many art forms that has been transformed by the digital revolution in the 90s, impacting not only the way people compose music but also the way people consume it. Nowadays, the most commonly used format for recording, processing, composing, distributing, and consuming music is digital. Online music platforms provide easy access to extensive music collections, underlining the importance of one of the most popular and widely used information filtering systems in the industry, the recommendation systems. Music recommendation systems allow for the exploration of large and complex collections of music, suggesting appropriate songs to a user that does not necessarily know all the available alternatives. The field of music information retrieval serves as a powerful knowledge resource for music recommendation systems. Its techniques rooted in psychoacoustics and signal processing empower their effectiveness of such systems.

1.4.1 Digital Music

Digitization with pulse-code modulation (PCM) allows for capturing and transforming the acoustic wave into an analogous digital representation that is easy to store and transmit [92]. During PCM, the amount of pressure change in the medium's particles relative to a baseline equilibrium is sampled at equally spaced intervals. A microphone, for example, is capable of converting the acoustic signal into an electrical signal utilizing its diaphragm that is sensitive to variations in air pressure. The disturbances in the membrane generate differences in the battery-powered voltage levels over time that can be recorded and stored as a digital signal. At a sample rate of 44.1 kHz, 44100 amplitude levels are sampled from the sound wave in the interval of a second, a long-time standard in high-quality music [9]. With this sample rate, a digital representation of a three minutes song approximates a total of eight million samples. Furthermore, each audio sample has a bit resolution, which is the amount of information about air pressure that it can carry. This means that in a 16-bit sample depth, the displacement of the membrane can be measured into 65536 different levels.

The digital format facilitates the storage and distribution of music, in such a way that both personal and public collections of music can grow enormously. The largest dance music store Beatport has over 7 million songs in its collection, while the world's most popular streaming service Spotify claims that 6000 songs are uploaded to the platform daily. Due to the extensive size of such collections, navigation and searching can be overwhelming for the users. Usually, listeners rely on browsing through metadata information (e.g. genre, name of the song, artist) linked to the digitized audio in order to find the desired song. Nevertheless, information about most of the songs is not known by the listener, showing that some music that could be interesting to the user remains undiscovered due to lack of prior knowledge.

1.4.2 Recommendation approaches

The two main approaches for designing recommendation systems are collaborative filtering and content-based filtering [89, 82]. Collaborative filtering recommendations aim to predict the interest of a user based on other users with similar musical taste [99, 116, 103, 94]. With this approach, if user A prefers ten songs from a collection and user B prefers nine of these songs but not song x, then the system will likely recommend song x to user B. The other approach to recommendation systems, content-based filtering, adopts a more descriptive representation of the songs, and user taste is defined in terms of item features [6, 108, 64]. With this approach, if a user prefers a large number of instrumental music from the collection, then the recommendation for that user will likely be an instrumental song. Moreover, some music recommendation systems have suggested an implementation that incorporates aspects of the two approaches in a variety of hybrid systems [10, 115].

It is important to notice that, although widely implemented, both approaches to music recommendation systems depend on existing recordings of users' preferences to exercise its function properly. Collaborative filtering needs multiple ratings from a great number of users. Furthermore, songs that are not rated by any user (e.g. new songs added to the system or unpopular songs) are never recommended, hindering the discovery of possibly relevant items through recommendations. Content-based filtering requires, at least, various ratings from a single user. Whenever a new user joins the system or in the case a new recommendation system is being developed from scratch, the history of user ratings is not readily available. This difficulty faced by new recommendation systems and new users is known as the cold-start problem. Luckily, a less noted approach offers a workaround to the cold-start problem, knowledge-based recommendation systems [24, 7]. Such approach is solely based on knowledge about their users, without the need for existing records of preference. The discovery of new songs is rooted in understanding the needs of the user and finding songs that meet these requirements.

1.4.3 Case-based reasoning

A prominent problem-solving methodology used in the knowledge-based approach to recommendation systems is case-based reasoning [56, 50]. The idea behind this methodology is to resolve a problem based on cases that were previously a successful solution. In the realm of music recommendation, the problem of finding songs that are appropriate to the user can be tackled with case-based reasoning. If we assume that users have a prior preference for one song, we can employ this song as an example of a successful solution to the search problem. Providing an exemplar to the system to retrieve similar items is known as query-by-example, and it has been one of the core implementations in case-based reasoning [41]. With the implementation of query-by-example, it is possible to recommend an ordered list of appropriate songs for being similar to the query. For that reason, a similarity measure that reflects the extent to which songs are alike is necessary for the query-by-example technique.

1.4.4 Music similarity

Assessing the object similarity between two pieces of music has been a prominent focus of the field of music information retrieval and it is often based on audio feature extraction [58, 73, 47]. Digital signal processing techniques provide a way to transform the audio data into representative features that are intended to better human interpretations. It is assumed that retrieving a song that has similar features to the query's features is an objective measure of music similarity. Audio features fall under one of three levels of abstraction, namely low-level, mid-level and high-level features [54]. Low-level representations are statistical descriptors resulting from a direct computation over the sample values in the audio signal. From a human interpretation perspective, these features are generally considered to be the farthest from conveying semantics. On the other hand, high-level features describe musical concepts more directly associated with human perception. Such features are often in the form of annotations that describe the song by their genre, their instrumentation, the emotions they convey, among others. Occupying an intermediate position, mid-level features apply a psychoacoustically informed transformation to the raw audio, such as to obtain a more meaningful description of the frequency spectrum.

One particular mid-level feature that has been used extensively in the literature concerning music similarity and music recommendation is the Mel-Frequency Cepstral Coefficients (MFCCs) [104, 22, 105]. MFCCs are a non-linear transformation of the raw digital audio. They are based on the human perception of pitch distances, which encompass sensitivity to changes in pitch at the lower frequencies. Regarding the primary musical elements, the general view is that MFCCs are a valid descriptor of the music's timbre and melodic content. Many similarity measures in music recommendation systems have implemented melody and timbre similarity as their main measure for music similarity. MFCCs are fundamental in most state-of-the-art approaches, weighing the importance of timbre information in music recommendation systems. Other descriptors are often used in conjunction with the MFCCs to incorporate information beyond melody and timbre in the system's similarity measure [123, 97, 5]. Similarity measures based on rhythmic mid-level representations have been proposed in the literature [86, 62] and also implemented in such multi-feature models [80, 57, 13]. Yet, pure rhythm similarity is rarely used in music recommendation, even though it has been realized in other similar applications [90, 81].

1.5 Problem description and research question

Researches show that music perception and music appreciation remain a challenge for CI users. Still, CI users consider music the second main acoustic stimulus [29] and a considerable amount of CI users report listening to three or more hours of music weekly [60]. Helping CI users to discover songs that meet their needs can be interesting for a better music listening experience. Up to this date, no music recommendation system that objectively considers the particularities of cochlear implant users has been developed. CI users can only rely on available music recommendation systems that consider in their retrieval poorly perceived elements of music, especially timbre and melody, and such systems might not provide appropriate recommendations for CI users. Therefore, the driving question behind this paper is the following: how can we develop a music recommendation system that is suitable for cochlear implant users?

2 Methods

This paper proposes a model for a case-based music recommendation considering the most prominent characteristics of the CI users' music listening experience. The model employs knowledge about CI users' music perception and music appreciation in its implementation of the case-based reasoning methodology. The first step consists of transforming songs from a music database into audio features that describe their rhythmic content. In that manner, the system implements the musical element most accurately perceived by CI users. Next, the user queries the system with a preferred song. Assuming familiarity with the query, the model considers the user's appreciation for familiar songs. In the following step, the system uses a distance measure between the query's and all other songs' rhythmic descriptors to find the most rhythmically similar songs. This step reflects the ability of the CI user to recognize rhythm similarity. The songs retrieved by similarity are re-ranked according to the tags that indicate semantic elements present in them. Songs with the highest scores in this new rank contain more musical elements that are better perceived or better appreciated by CI users. For example, songs with a tag indicating the presence of vocals receive one point for it in their re-ranking score. A general description of the model is portrayed in a scheme (Figure 1).

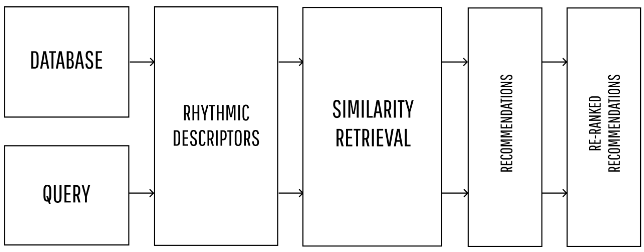

Figure 1: An overview of the case-based music recommendation system for CI users.

2.1 The dataset

The chosen music database for implementing the model was the MagnaTagATune Dataset [61]. It contains 25854 raw audio clips with a sample rate of 16kHz and a total duration of approximately 29.1 seconds. Songs are cut at different positions to create multiple audio clips without overlap. Audio metadata is available, such as the name of the song, artist, and album. The dataset also includes high-level features in the form of tags symbolizing elements in the songs (Figure 2). The annotations consist of binary values indicating instrumentation (e.g. guitar, piano, violin, male vocal), genre (e.g. techno, classic, opera), speed (e.g. fast, slow), or loudness (e.g. loud, soft, of a song). A game platform called TagATune collected the tags, in which two players were either given the same or different audio clips and needed to agree on which elements were present in the clips.

It is worth noting that the original full songs from which the audio clips were extracted pertain to different genres. For that reason, the audio clips are expected to represent a wide variety of rhythms, melodies, and timbres, according to the musical style in which they were composed. Real-life applications with more extensive databases can have a much larger number of genres. Yet, the crucial notion is that the dataset can capture the variability of musical characteristics and primary elements in a collection of songs. As a pre-processing step, tags that are synonyms were grouped, considering the minimal disparities between annotations such as “beat” and “beats”, “no singer” and “no singing”, or ”vocal” and ”singer”.

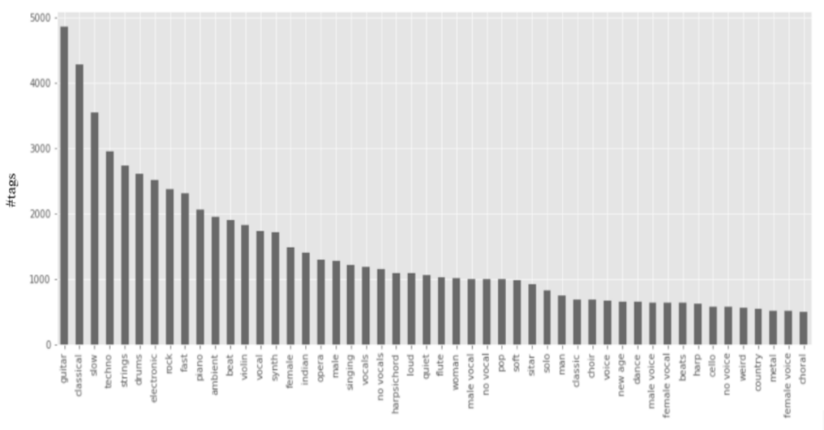

Figure 2: Number of annotations per tag before aggregation by synonyms

2.2 Audio features

In the next step, audio data were transformed from raw digital audio into mid-level audio features by extracting their respective rhythmic descriptors. For that, the Python library *rp_extract* was used to process the array of samples representing each audio clip and extract both the *Rhythm Patterns* and *Rhythm Histograms* features, as proposed in [62]. *Rhythm Patterns* represent rhythmic structure over a range of frequency bands, whereas *Rhythm Histograms* describes the overall rhythmic characteristics of a music piece. Their extraction has multiple steps, employing digital signal processing techniques and accounting for psychoacoustic knowledge about human psychological responses associated with sound. Researchers in [62] experimented with a variety of techniques in the extraction process of both features to assess how their combination impacts accuracy in music similarity tasks. The specific method depicted in Figure 3 and the use of these two rhythmic descriptors for retrieval in three other datasets (ISMIRhythm, ISMIRGenre, and GTZAN) yielded promising results.

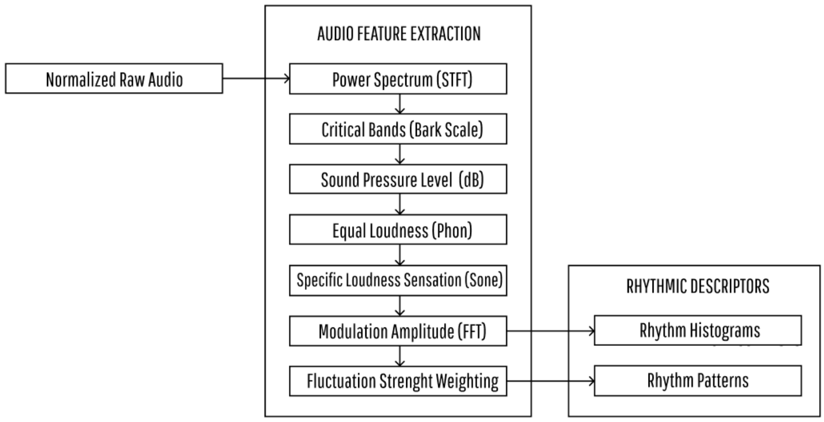

Figure 3: Steps in feature extraction process: from normalized raw audio to Rhythm Patterns and Rhythm Histograms

First, for each song in the dataset, six-second excerpts were taken from the raw audio data, and the following process was done on all the excerpts in separation. The STFT (short-time Fourier transform) was calculated to dissect the audio data into its distribution of power in the frequency spectrum, using a Hanning window with a window size of 23ms and a 50% overlap. Next, frequency powers of similar frequencies were grouped in 24 critical bands called *Barks*, stemming from the idea that the range of audible frequency can be subdivided for simplification [124]. The following step accounted for the auditory masking effect by applying a spreading function over the 24 critical bands to reckon their interference with each other [95]. Then, the audio samples were converted from pressure values (Pascal) to decibel (dB), a logarithmic unit to measure relative sound levels referent to perceived intensity [107]. In the subsequent step, loudness levels were converted to Phon units, to incorporate the levels of perceived equal loudness, considering that the perceived loudness of sounds with the same decibel level depends on their frequency [107]. Samples in Phone were further converted into Sone units to achieve specific loudness sensations, considering that levels of perceived equal loudness differ according to the sound levels.

After the transformation into Sone units, the data representation consisted of specific loudness sensations over time for the 24 critical bands. A fast Fourier transform (FFT) was further applied to obtain a time-invariant description of the amplitude modulations in each band. Considering that amplitude modulations between 0.2 Hz and 10 Hz are perceived as rhythm, only this range of the obtained representation was kept (from 12 to 600 BPM). The first rhythm descriptor, the *Rhythm Histograms*, was derived from this representation. The frequency range from 0.2 Hz to 10 Hz is divided into 60 bins and all the magnitudes in the critical bands belonging to the same frequency bin were summed. The *Rhythm Histograms* from all the six-second excerpts of the same song were ultimately aggregated by calculating their median over the resulting representations. The result was a histogram that describes the general magnitude concentrated in the different bands indicating tempo estimations. In order to compute the other rhythm descriptor, the *Rhythm Patterns*, the amplitude modulations in each band after the FFT were weighted based on the human perception of the strength of these fluctuations. Fluctuation strength perception is maximal around 4 Hz [23]. Finally, the *Rhythm Patterns* from all the excerpts of the same song were also aggregated to their median. This resulted in one single time-invariant representation of all the songs in the dataset with respect to their rhythmic descriptors.

2.3 Similarity retrieval and re-ranking strategy

With every song converted into their rhythmic descriptors, a new feature space was conceived, allowing for audio comparison concerning rhythm. Before calculating the distance between songs, *Rhythm Patterns* and *Rhythm Histograms* from each song were standardized and concatenated. In the following step, the Euclidean distance in the feature space was calculated - from the query features to every other song's features - resulting in similarity measurements. Then, the dataset was sorted in a way that the five most rhythmically similar songs were on top and retrieved in an ordered list. A new matrix was further created, solely containing tags for elements that are better perceived or appreciated by CI users. The chosen tags were the following: "singer" and "rap", for containing the highly estimated lyrical content, "drums" and "beat", for containing very sharp attacks, "piano" for being the most well-appreciated instrument, and "slow" for the CI user's appreciation for slow rhythms.

Some music recommendation systems have employed a strategy to re-rank the retrieved songs before giving a final recommendation [37, 122, 49]. The x recommendations retrieved by similarity proceed to an algorithm that gives them new scores based on specific heuristics depending on the context. As a final step in the model, for each of the five retrieved songs, the summation of tag values yielded a new score for re-ranking. The new highest-scoring songs were relocated to the top of the recommendation list. In case there was the same score for two songs, the most similar remained above the other.

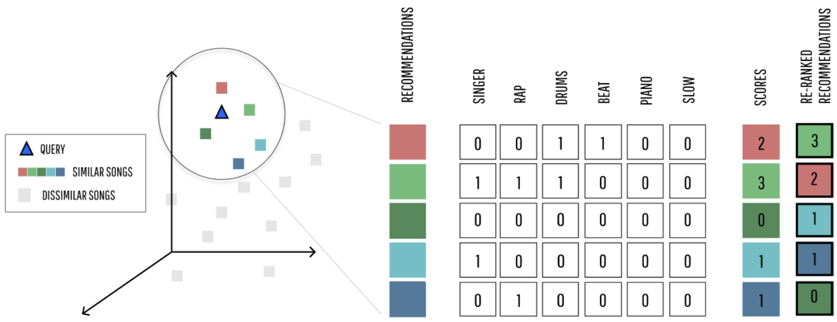

Figure 4: Similarity retrieval of five songs by closeness in a feature space and re-ranking according to annotations

3 Results

The first step of the model proposed to extract from the raw audio a representation of the rhythmic descriptors, the Rhythm Patterns (Figure 5) and Rhythm Histograms (Figure 6). Five songs from different music genres were picked arbitrarily to illustrate the feature extraction performed by the system, allowing for a comparison between their waveform and rhythmic descriptors.

It was possible to observe that some of these songs, such as "Grayson Wray - In 1671", have more salient peaks in the waveform than the other. By listening to the audio file, it was possible to note that those peaks arise from a strong drum beat pattern with a piano. After nine seconds, when the waveform becomes fuller, a male singing

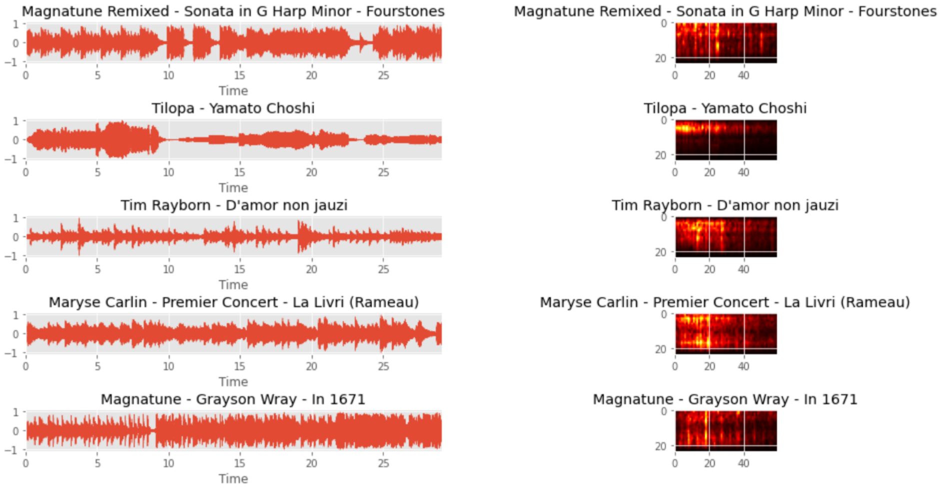

Figure 5: Five songs in the raw form (left) and their respective Rhythm Patterns (right).

voice starts. The marked percussive rhythm pattern was reflected as a greater intensity in its Rhythm Patterns. A white-colored column was localized on the 18th bin across multiple bands, indicating the modulation frequency of 3 Hz or 180 BPM. On the Rhythm Histogram, this was also observed as a salient 18th bin. Another song, "Tilopa - Yamato Choshi", consists of a single flute played without any other instrument or salient demarcation of rhythm. Consequently, the waveform was the most plane among the five songs. The Rhythm Patterns representation had less localized in terms of modulation frequency bins. However, it was possible to observe intensity spread horizontally around the extension of band 5, indicating less rhythmic content. This was also reflected on the Rhythm Histograms without any particular salience (Figure 7).

To simulate the preference of a CI user, two songs were given as a query to the system to retrieve other similar songs. It is possible to observe that the retrieved songs shared similar rhythmic patterns with its respective query (Figure 8). In the case of songs that were similar to "Grayson Wray - In 1671", for instance, a noticeable similarity in the intensity of specific amplitude modulations could be observed. On the other hands, songs that were similar to "Tilopa - Yamato Choshi" also had no apparent salient bin as the query. A comparison of Rhythm Histograms did not provide any further insights besides these and was omitted here. Finally, to explore the re-ranking strategy, the five songs retrieved with the query "Grayson Wray - In 1671" proceeded to the scoring step. In Figure 9, a detailed scoring for each song is showed. After the strategy, the song "Tom Paul - It's an easy life" was pushed to the top of the recommendations.

4 Discussion

The model proposed by this paper is based on studies that assessed the music listening experience of CI users, even though there was wide variability among CI users regarding both music appreciation and music perception levels. Some influential factors are musical background, age at implantation, CI's sound processing strategies, and whether users received the implants before or after learning their primary language. For instance, most studies focus on adult CI users who acquired their first language before the onset of their hearing condition. Also, some CI users simply do not enjoy music at all, which shows that the system could only help if there is a preexisting level of appre-

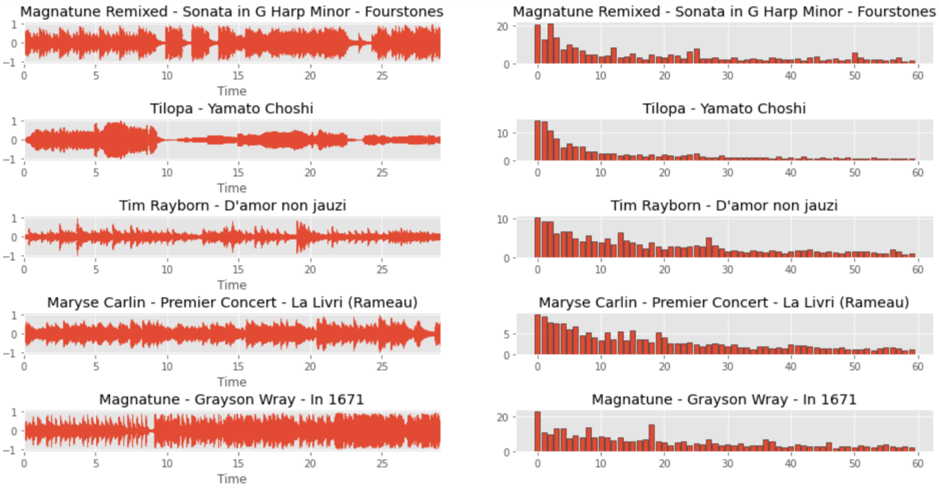

Figure 6: Five songs in the raw form (left) and their respective Rhythm Histograms (right).

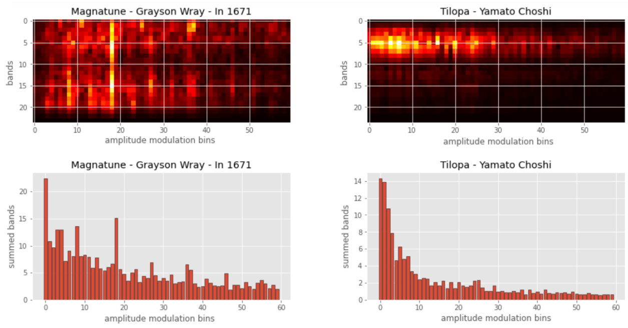

Figure 7: Comparison of "Grayson Wray - In 1671" (left) and "Tilopa - Yamato Choshi" (right) with respect to their rhythmic descriptors.

ciation, at least for the queried song. Studies also use relatively simplified stimulus (e.g. pure tones) on perception tasks (besides recognition of music excerpts), which is not the reality of songs. For that reason, the findings of these studies had to be generalized before being implemented in the system.

The feature extraction process allowed for the transformation of a music database and a query into rhythmic descriptors that served as a space for comparison. The song "Tom Paul - It's an easy life" was retrieved as the most similar to the query "Grayson Wray - In 1671" and it contained a similar beat pattern, as judged by the author of this paper. However, a more extensive evaluation of the results with human participants, especially with CI users, is indispensable for the judgment rhythm similarity. Another noticeable element of similarity retrieval with this dataset is that the same song was retrieved twice with the query "Grayson Wray - In 1671". This repetition results from the presence of multiple audio clips from the same songs in the dataset. Considering that users will listen to the full song, a conversion method that filters the duplicates is needed. Furthermore, other rhythmic descriptors have been proposed [42] and could also be a future incorporation to the system. Furthermore, because the processes applied in the extraction of *Rhythm Patterns* and *Rhythm Histograms* are based on psychoacoustic measures of normal-hearing people, the development of new CI-based feature extraction

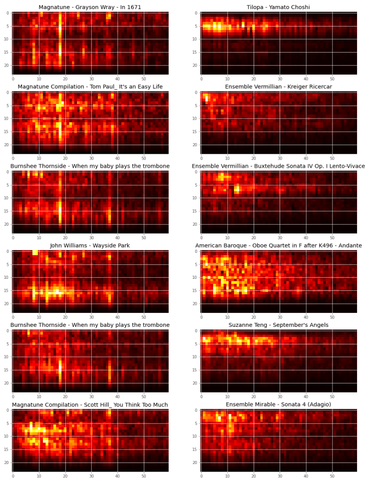

Figure 8: The query "Grayson Wray - In 1671" (top left) and the query "Tilopa - Yamato Choshi" (top right) with an ordered list of most rhythmically similar songs bellow them.

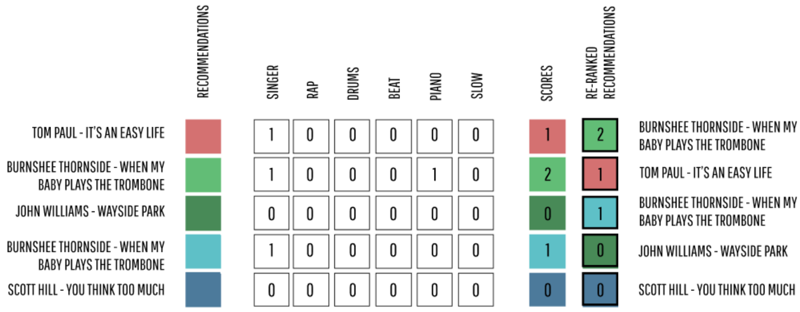

Figure 9: The most rhythmically similar recommendations retrieved with the query "Grayson Wray - In 1671" re-ranked according to the tags

methods is an interesting area of investigation.

Concerning the use of tags in the re-ranking strategy, a scalability issue arises due to its reliance on human annotations. Methods for automatic song tagging are encouraged [106, 88]. Still, it is also not guaranteed that such tags can be reliable, considering that annotations are based on subjective human judgments. For instance, even though the song "Tom Paul - It's an easy life" has a prominent drum pattern, it did not have a "drum" annotation linked to it.

Finally, other design extensions could be an attractive addition to the system. For instance, a processing scheme for enhancing vocals and percussions via sound source separation can have a positive impact on the CI users' music experience [12, 11, 87]. Also, musical elements that were not considered in the model can have an impact on the music listening experience, such as music compression [118] and amount of reverberation [93, 16]. Furthermore, being able to read and follow the lyrics while listening is linked to appraisal, which could be a possible addition to the system [67]. Finally, some songs may remain undiscovered for never being among the five most similar songs to any query, a limitation found in recommendation systems based on audio content [26].

5 Conclusion

This paper proposed a case-based music recommendation system that objectively considered aspects of the CI user's music experience. It examined studies about music perception and music appreciation of the users as a foundation for the design choices. The model recommends rhythmically similar songs, considering that CI users can perceived rhythm similarity. These songs were also re-ranked according to the presence of elements that are appreciated or well-perceived. Final recommendations are not only rhythmically similar, but also more likely to have musical elements with a positive influence in the CI users' music listening experience. This model is an encouraging starting point for recommendation systems suitable for cochlear implant users, even though future considerations about cochlear implants, music recommendation systems, and psychoacoustics can pave the way for a more fitting system.

References

[1] Chapter 7 sound propagation in the sea. In Jerald W. Caruthers, editor, *Fundamentals of Marine Acoustics*, volume 18 of *Elsevier Oceanography Series*, pages 99–112. Elsevier, 1977.

[2] Paul J. Abbas, Michelle L. Hughes, Carolyn J. Brown, Charles A. Miller, and Heather South. Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential. *Audiology and Neuro-Otology*, 9(4):203–213, 2004.

[3] Jeremy F. Alm and James S. Walker. Time-frequency analysis of musical instruments. *SIAM Review*, 44(3):457–476, 2002.

[4] Keith Attenborough. *Sound Propagation in the Atmosphere*, pages 113–147. Springer New York, New York, NY, 2007.

[5] D. Bogdanov, J. Serra, N. Wack, P. Herrera, and X. Serra. Unifying low-level and high-level music similarity measures. *IEEE Transactions on Multimedia*, 13(4):687–701, August 2011.

[6] Dmitry Bogdanov, Martín Haro, Ferdinand Fuhrmann, Anna Xambó, Emilia Gómez, and Perfecto Herrera. Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing & Management*, 49(1):13–33, January 2013.

[7] Sarah Bouraga, Ivan Jureta, Stéphane Faulkner, and Caroline Herssens. Knowledge-based recommendation systems. *International Journal of Intelligent Information Technologies*, 10(2):1–19, April 2014.

[8] Casady Bowman and Takashi Yamauchi. Perceiving categorical emotion in sound: The role of timbre. *Psychomusicology: Music, Mind, and Brain*, 26(1):15–25, 2016.

[9] K. Brandenburg and M. Bosi. Overview of mpeg audio : Current and future standards for low-bit-rate audio coding. *Journal of The Audio Engineering Society*, 45:4–21, 1997.

[10] Poonam B.Thorat, R. M. Goudar, and Sunita Barve. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4):31–36, January 2015.

[11] Wim Buyens, Bas van Dijk, Marc Moonen, and Jan Wouters. Music mixing preferences of cochlear implant recipients: A pilot study. *International Journal of Audiology*, 53(5):294–301, January 2014.

[12] Wim Buyens, Bas van Dijk, Jan Wouters, and Marc Moonen. A harmonic/percussive sound separation based music pre-processing scheme for cochlear implant users. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5, 2013.

[13] Pedro Cano, Markus Koppenberger, and Nicolas Wack. An industrial-strength content-based music recommendation system. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05*. ACM Press, 2005.

[14] Peter Cariani. Temporal coding of periodicity pitch in the auditory system: An overview. *Neural Plasticity*, 6(4):147–172, 1999.

[15] Wendy A. Castleman and Randy L. Diehl. Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics*, 24(4):383–398, 1996.

[16] Michael V. Certo, Gavriel D. Kohlberg, Divya A. Chari, Dean M. Mancuso, and Anil K. Lalwani. Reverberation time influences musical enjoyment with cochlear implants. *Otology & Neurotology*, 36(2):e46–e50, February 2015.

[17] Graeme Clark. Cochlear implants. In *Speech Processing in the Auditory System*, pages 422–462. Springer-Verlag.

[18] Patrick J. Donnelly, Benjamin Z. Guo, and Charles J. Limb. Perceptual fusion of polyphonic pitch in cochlear implant users. *The Journal of the Acoustical Society of America*, 126(5):EL128–EL133, November 2009.

[19] Ward. R. Drennan, Jacob J. Oleson, Kate Gfeller, Jillian Crosson, Virginia D. Driscoll, Jong Ho Won, Elizabeth S. Anderson, and Jay T. Rubinstein. Clinical evaluation of music perception, appraisal and experience in cochlear implant users. *International Journal of Audiology*, 54(2):114–123, February 2015.

[20] Giorgos Dritsakis, Rachel M. van Besouw, and Aoife O' Meara. Impact of music on the quality of life of cochlear implant users: a focus group study. *Cochlear Implants International*, 18(4):207–215, 2017. PMID: 28326996.

[21] Albert SB Edge and Zheng-Yi Chen. Hair cell regeneration. *Current Opinion in Neurobiology*, 18(4):377–382, 2008. Sensory systems.

[22] Daniel P. W. Ellis. Classifying music audio with timbral and chroma features. 2007.

[23] Hugo Fastl. Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise. *Hearing Research*, 8(1):59–69, September 1982.

[24] Alexander Felfernig, Michael Jeran, Gerald Ninaus, Florian Reinfrank, Stefan Reiterer, and Martin Stettinger. Basic approaches in recommendation systems. In *Recommendation Systems in Software Engineering*, pages 15–37. Springer Berlin Heidelberg, December 2013.

[25] Bruce L. Fetterman and Elizabeth H. Domico. Speech recognition in background noise of cochlear implant patients. *Otolaryngology - Head and Neck Surgery*, 126(3):257–263, 2002.

[26] Arthur Flexer, Martin Gasser, and Dominik Schnitzer. Limitations of interactive music recommendation based on audio content. In *Proceedings of the 5th Audio Mostly Conference on A Conference on Interaction with Sound - AM '10*. ACM Press, 2010.

[27] Paul Fraisse. Time and rhythm perception. In *Perceptual Coding*, pages 203–254. Elsevier, 1978.

[28] K. Gfeller, A. Christ, K. John, S. Witt, and M. Mehr. The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal hearing adults. *Journal of Music Therapy*, 40(2):78–112, June 2003.

[29] K. Gfeller, A. Christ, J. Knutson, S. Witt, K. Murray, and R. Tyler. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology*, 11 7:390–406, 2000.

[30] Kate Gfeller and Charissa R. Lansing. Melodic, rhythmic, and timbral perception of adult cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 34(4):916–920, August 1991.

[31] Kate Gfeller, Carol Olszewski, Marly Rychener, Kimberly Sena, John F. Knutson, Shelley Witt, and Beth Macpherson. Recognition of ???real-world??? musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear and Hearing*, 26(3):237–250, June 2005.

[32] Kate Gfeller, Christopher Turner, Maureen Mehr, George Woodworth, Robert Fearn, John F Knutson, Shelley Witt, and Julie Stordahl. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International*, 3(1):29–53, March 2002.

[33] Kate Gfeller, Christopher Turner, Jacob Oleson, Xuyang Zhang, Bruce Gantz, Rebecca Froman, and Carol Olszewski. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear & Hearing*, 28(3):412–423, Jun 2007.

[34] Kate Gfeller, Shelley Witt, Maureen A. Mehr, George Woodworth, and John Knutson. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology & Laryngology*, 111(4):349–356, April 2002.

[35] Kate Gfeller, George Woodworth, Donald A. Robin, Shelley Witt, and John F. Knutson. Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and Hearing*, 18(3):252–260, June 1997.

[36] Julius L. Goldstein. Changing roles in the cochlea: Bandpass filtering by the organ of corti and additive amplification on the basilar membrane. *The Journal of the Acoustical Society of America*, 92(4):2407–2407, October 1992.

[37] Boning Gong, Mesut Kaya, and Nava Tintarev. Contextual personalized re-ranking of music recommendations through audio features, 2020.

[38] Juan D. Goutman, A. Belén Elgoyhen, and María Eugenia Gómez-Casati. Cochlear hair cells: The sound-sensing machines. *FEBS Letters*, 589(22):3354–3361, 2015. South America Special Issue.

[39] Donald D. Greenwood. Auditory masking and the critical band. *The Journal of the Acoustical Society of America*, 33(4):484–502, April 1961.

[40] Julia C. Hailstone, Rohani Omar, Susie M. D. Henley, Chris Frost, Michael G. Kenward, and Jason D. Warren. It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, 62(11):2141–2155, November 2009.

[41] H. HARB and L. CHEN. A QUERY BY EXAMPLE MUSIC RETRIEVAL ALGORITHM. In *Digital Media Processing for Multimedia Interactive Services*. WORLD SCIENTIFIC, March 2003.

[42] André Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1):176–185, January 2011.

[43] Yoshiyuki Horii. Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research*, 18(1):192–201, 1975.

[44] William F. House. Cochlear implants. *Annals of Otology, Rhinology & Laryngology*, 85(3_suppl):3–3, 1976.

[45] Ying Hsiao, Valeriy Shafiro, Chad Walker, Jasper Oh, Megan Hebb, Kelly Brown, Stan Sheft, Kara Vasil, and Aaron Moberly. Perception of musical instruments and music genres in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 145:1876–1876, 03 2019.

[46] Colin Humphries, Einat Liebenthal, and Jeffrey R. Binder. Tonotopic organization of human auditory cortex. *NeuroImage*, 50(3):1202–1211, 2010.

[47] Jesper HØjvang Jensen, Mads GrÆsbØll Christensen, Daniel P. W. Ellis, and SØren Holdt Jensen. Quantitative analysis of a common audio similarity measure. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(4):693–703, May 2009.

[48] Patrik N. Juslin. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press, July 1993.

[49] Surya Kallumadi, Bhaskar Mitra, and Tereza Iofciu. A line in the sand: Recommendation or ad-hoc retrieval?, 2018.

[50] Karen Ketler. Case-based reasoning: An introduction. *Expert Systems with Applications*, 6(1):3–8, January 1993.

[51] Irene Kim, Eunice Yang, Patrick J. Donnelly, and Charles J. Limb. Preservation of rhythmic clocking in cochlear implant users: A study of isochronous versus anisochronous beat detection. *Trends in Amplification*, 14(3):164–169, September 2010.

[52] T. Kitahara, M. Goto, and H.G. Okuno. Musical instrument identification based on f0-dependent multivariate normal distribution. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, volume 5, pages V–421, 2003.

[53] Tetsuro Kitahara, Masataka Goto, and Hiroshi G. Okuno. Pitch-dependent identification of musical instrument sounds. *Applied Intelligence*, 23(3):267–275, December 2005.

[54] Peter Knees and Markus Schedl. *Music Similarity and Retrieval*. Springer Berlin Heidelberg, 2016.

[55] Stefan Koelsch and Walter A. Siebel. Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9(12):578–584, 2005.

[56] Janet L. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6(1):3–34, 1992.

[57] Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, and Suh-Yin Lee. Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05*. ACM Press, 2005.

[58] Frank Kurth and Meinard Muller. Efficient index-based audio matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):382–395, February 2008.

[59] J. M. Lafferty and L. Rubin. Foundations of vacuum science and technology. *Physics Today*, 52:86–88, 1999.

[60] Luis Lassaletta, Alejandro Castro, Marta Bastarrica, Rosa Pérez-Mora, Rosario Madero, Josefa De Sarriá, and Javier Gavilán. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica*, 127(7):682–686, January 2007.

[61] Edith Law, Kris West, Michael Mandel, Mert Bay, and J. Downie. Evaluation of algorithms using games: The case of music tagging. pages 387–392, 01 2009.

[62] Thomas Lidy and Andreas Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. pages 34–41, 01 2005.

[63] Charles J. Limb. Structural and functional neural correlates of music perception. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology*, 288A(4):435–446, 2006.

[64] Ning-Han Liu. Comparison of content-based music recommendation using different distance estimation methods. *Applied Intelligence*, 38(2):160–174, June 2012.

[65] P.C. Loizou. Introduction to cochlear implants. *IEEE Engineering in Medicine and Biology Magazine*, 18(1):32–42, 1999.

[66] Valerie Looi, Hugh McDermott, Colette McKay, and Louise Hickson. Pitch discrimination and melody recognition by cochlear implant users. *International Congress Series*, 1273:197–200, 2004. Cochlear Implants. Proceedings of the VIII International Cochlear Implant Conference.

[67] Valerie Looi and Jennifer She. Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. *International Journal of Audiology*, 49(2):116–128, January 2010.

[68] MD Luis Lassaletta, MD Alejandro Castro, Marta Bastarrica, MD Rosa Pérez-Mora, MD Belén Herrán, MD Lorena Sanz, MD M. Josefa de Sarriá, and MD Javier Gavilán. Changes in listening habits and quality of musical sound after cochlear implantation. *Otolaryngology–Head and Neck Surgery*, 138(3):363–367, 2008. PMID: 18312886.

[69] Robert A. Lutfi. Human sound source identification. In *Auditory Perception of Sound Sources*, pages 13–42. Springer US.

[70] Olivier Macherey and Alexia Delpierre. Perception of musical timbre by cochlear implant listeners. *Ear & Hearing*, 34(4):426–436, July 2013.

[71] Jeremy Marozeau and Wiebke Lamping. Timbre perception with cochlear implants. In *Timbre: Acoustics, Perception, and Cognition*, pages 273–293. Springer International Publishing, 2019.

[72] Josh H McDermott and Andrew J Oxenham. Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, 18(4):452–463, 2008. Sensory systems.

[73] Brian McFee, Luke Barrington, and Gert Lanckriet. Learning content similarity for music recommendation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(8):2207–2218, October 2012.

[74] Lela Migirov, Jona Kronenberg, and Yael Henkin. Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology & Laryngology*, 118(5):350–355, May 2009.

[75] S Mirza, SA Douglas, P Lindsey, T Hildreth, and M Hawthorne. Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear Implants International*, 4(2):85–95, June 2003.

[76] Michelle Moran, Alexandra Rousset, and Valerie Looi. Music appreciation and music listening in prelingual and postlingually deaf adult cochlear implant recipients. *International Journal of Audiology*, 55(sup2):S57–S63, April 2016.

[77] Peggy B. Nelson and Su-Hyun Jin. Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 115(5):2286–2294, May 2004.

[78] Sirpa Nummela and J G M Thewissen. The physics of sound in air and water. *Sensory Evolution on the Threshold: Adaptations in Secondarily Aquatic Vertebrates*, pages 175–182, 02 2008.

[79] Andrew J. Oxenham. Pitch perception. *Journal of Neuroscience*, 32(39):13335–13338, 2012.

[80] Elias Pampalk. Audio-based music similarity and retrieval: Combining a spectral similarity model with information extracted from fluctuation patterns. implementation submitted to the 3rd annual music information retrieval evaluation exchange. In *in Proceedings of the International Symposium on Music Information Retrieval*, 2006.

[81] Elias Pampalk, Andreas Rauber, and Dieter Merkl. Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia - MULTIMEDIA 02*. ACM Press, 2002.

[82] Michael J. Pazzani and Daniel Billsus. Content-based recommendation systems. In *The Adaptive Web*, pages 325–341. Springer Berlin Heidelberg, 2007.

[83] Gamaliel Percino, Peter Klimek, and Stefan Thurner. Instrumentational complexity of music genres and why simplicity sells. *PLoS ONE*, 9(12):e115255, December 2014.

[84] Jessica Phillips-Silver, Petri Toiviainen, Nathalie Gosselin, Christine Turgeon, Franco Lepore, and Isabelle Peretz. Cochlear implant users move in time to the beat of drum music. *Hearing Research*, 321:25–34, March 2015.

[85] Patrick N. Plyler, Junghwa Bahng, and Deborah von Hapsburg. The acceptance of background noise in adult cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 51(2):502–515, April 2008.

[86] Tim Pohle, Dominik Schnitzer, Markus Schedl, Peter Knees, and Gerhard Widmer. On rhythm and general music similarity. pages 525–530, 01 2009.

[87] Jordi Pons, Jordi Janer, Thilo Rode, and Waldo Nogueira. Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, 140(6):4338–4349, December 2016.

[88] Jordi Pons and Xavier Serra. musicnn: Pre-trained convolutional neural networks for music audio tagging. *CoRR*, abs/1909.06654, 2019.

[89] Sandeep K. Raghuwanshi and R. K. Pateriya. Collaborative filtering techniques in recommendation systems. In *Data, Engineering and Applications*, pages 11–21. Springer Singapore, 2019.

[90] Andreas Rauber, Elias Pampalk, and Dieter Merkl. The SOM-enhanced JukeBox: Organization and visualization of music collections based on perceptual models. *Journal of New Music Research*, 32(2):193–210, June 2003.

[91] GL Romani, SJ Williamson, and L Kaufman. Tonotopic organization of the human auditory cortex. *Science*, 216(4552):1339–1340, 1982.

[92] E.R. Rout and A.H. Jones. The use of pulse code modulation for point-to-point music transmission. *Radio and Electronic Engineer*, 37(4):199, 1969.

[93] Alexis T. Roy, Michelle Vigeant, Tina Munjal, Courtney Carver, Patpong Jiradejvong, and Charles J. Limb. Reverberation negatively impacts musical sound quality for cochlear implant users. *Cochlear Implants International*, 16(sup3):S105–S113, September 2015.

[94] Diego Sánchez-Moreno, Ana B. Gil González, M. Dolores Muñoz Vicente, Vivian F. López Batista, and María N. Moreno García. A collaborative filtering method for music recommendation using playing coefficients for artists and users. *Expert Systems with Applications*, 66:234–244, December 2016.

[95] M. R. Schroeder, B. S. Atal, and J. L. Hall. Optimizing digital speech coders by exploiting masking properties of the human ear. *The Journal of the Acoustical Society of America*, 66(6):1647–1652, December 1979.

[96] Lisa Schwartz. Understanding silence: Meaning and interpretation. *Performance Research*, 4(3):8–11, 1999.

[97] Bo Shao, M. Ogihara, Dingding Wang, and Tao Li. Music recommendation based on acoustic features and user access patterns. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(8):1602–1611, November 2009.

[98] Fernanda Soares, Nayara Freitas Fernandes, Marina Morettin, Orozimbo Alves Costa Filho, and Maria CecÃlia Bevilacqua. The relationship between the speech perception and the degree of satisfaction among adult users of cochlear implants. *International Archives of Otorhinolaryngology*, 17:202 – 207, 06 2013.

[99] Mohammad Soleymani, Anna Aljanaki, Frans Wiering, and Remco C. Veltkamp. Content-based music recommendation using underlying music preference structure. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, June 2015.

[100] William C. Stebbins, Joseph E. Hawkins, Lars-Göran Johnsson, and David B. Moody. Hearing thresholds with outer and inner hair cell loss. *American Journal of Otolaryngology*, 1(1):15–27, 1979.

[101] Mark J Steedman. The perception of musical rhythm and metre. *Perception*, 6(5):555–569, October 1977.

[102] Nikolaus Steinbeis, Stefan Koelsch, and John A. Sloboda. The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, 18(8):1380–1393, August 2006.

[103] M. Sunitha and T. Adilakshmi. Music recommendation system with user-based and item-based collaborative filtering technique. In *Networking Communication and Data Knowledge Engineering*, pages 267–278. Springer Singapore, November 2017.

[104] H. Terasawa, M. Slaney, and J. Berger. The thirteen colors of timbre. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.* IEEE.

[105] R. Thiruvengatanadhan, P. Dhanalakshmi, and S. Palanivel. GMM based indexing and retrieval of music using MFCC and MPEG-7 features. In *Advances in Intelligent Systems and Computing*, pages 363–370. Springer International Publishing, 2015.

[106] Derek Tingle, Youngmoo E. Kim, and Douglas Turnbull. Exploring automatic music annotation with "acoustically-objective" tags. In *Proceedings of the international conference on Multimedia information retrieval - MIR '10*. ACM Press, 2010.

[107] A. Tumarkin. The decibel, the phon and the sone. *The Journal of Laryngology & Otology*, 64(4):178–188, April 1950.

[108] Aäron van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In Christopher Burges, Léon Bottou, Max Welling, Zoubin Ghahramani, and Kilian Weinberger, editors, *Advances in Neural Information Processing Systems 26 (2013)*, volume 26, page 9. Neural Information Processing Systems Foundation (NIPS), 2013.

[109] Andrew E. Vandali, Catherine Sucher, David J. Tsang, Colette M. McKay, Jason W. D. Chew, and Hugh J. McDermott. Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *The Journal of the Acoustical Society of America*, 117(5):3126–3138, May 2005.

[110] Wuqing Wang, Ning Zhou, and Li Xu. Musical pitch and lexical tone perception with cochlear implants. *International journal of audiology*, 50:270–8, 12 2010.

[111] Richard M. Warren. 1 - sound and the auditory system. In Richard M. Warren, editor, *Auditory Perception*, volume 109 of *Pergamon General Psychology Series*, pages 1–30. Pergamon, 1982.

[112] C. Mark Wessinger, Michael H. Buonocore, Clif L. Kussmaul, and George R. Mangun. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, 5(1):18–25, 1997.

[113] Jan Wouters, Hugh Joseph McDermott, and Tom Francart. Sound coding in cochlear implants: From electric pulses to hearing. *IEEE Signal Processing Magazine*, 32(2):67–80, 2015.

[114] Rose Wright and Rosalie M. Uchanski. Music perception and appraisal: Cochlear implant users and simulated cochlear implant listening. *Journal of the American Academy of Audiology*, 23(05):350–365, May 2012.

[115] Yan Yan, Tianlong Liu, and Zhenyu Wang. A music recommendation algorithm based on hybrid collaborative filtering technique. In *Communications in Computer and Information Science*, pages 233–240. Springer Singapore, 2015.

[116] Billy Yapriady and Alexandra L. Uitdenbogerd. Combining demographic data with collaborative filtering for automatic music recommendation. In *Lecture Notes in Computer Science*, pages 201–207. Springer Berlin Heidelberg, 2005.

[117] Mustafa Yüksel, Atılım Atılgan, and Ayça Çiprut. Music listening habits and music perception abilities of prelingually deafened adolescent cochlear implant recipients. *Journal of the American Academy of Audiology*, 31(10):740–745, November 2020.

[118] Fan-Gand Zeng. Compression and cochlear implants. In *Compression: From Cochlea to Cochlear Implants*, pages 184–220. Springer-Verlag.

[119] Fan-Gang Zeng, Stephen Rebscher, William Harrison, Xiaoan Sun, and Haihong Feng. Cochlear implants: System design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, 1:115–142, 2008.

[120] Fan-Gang Zeng, Qing Tang, and Thomas Lu. Abnormal pitch perception produced by cochlear implant stimulation. *PLoS ONE*, 9(2):e88662, February 2014.

[121] Weixia Zhang, Fang Liu, Linshu Zhou, Wanqi Wang, Hanyuan Jiang, and Cunmei Jiang. The effects of timbre on neural responses to musical emotion. *Music Perception*, 37(2):134–146, December 2019.

[122] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. Auralist. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. ACM Press, 2012.

[123] Xuan Zhu, Yuan-Yuan Shi, Hyoung-Gook Kim, and Ki-Wan Eom. An integrated music recommendation system. *IEEE Transactions on Consumer Electronics*, 52(3):917–925, August 2006.

[124] E. Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2):248–248, February 1961.