Music, Mind and Technology - Assignment 1

Sreeja Guduri (2021102007)

1.1)
Summary of four abstracts:

a. **Predicting musical chill moments with inferential meta-analytics** (Dumas RE, Beck J, Prust A, Katz G, Moe PJ, El-Said D2, de Los Angeles C, Menon V, Levitin DJ)

Summary:
Chill moments or impactful moments refer to segments in a song that cause intense pleasure and physiological reactions similar to those associated with food, drugs and sex. Although some impactful moments could be distinct to a particular individual because of personal experiences, a lot of music produces chills in a large proportion of listeners (e.g. Albinoni "Adagio in G").

Thus, the main aim of the researchers was to discover musical segments that cause chills using an inferential software (called MIIR), and test the hypothesis that the software's predictions align with fMRI imaging of the brain and a behavioural study of participants' responses.

In the fMRI experiment, seventeen musically untrained participants were made to listen to 9 minute long symphonies by the composer William Boyce (1711–1779). The whole brain was scanned to determine where the activation levels were higher. In the behavioural study, participants listened to songs that they chose and ranked regions of the song based on the impact elicited. This data was then correlated with the output of the software.

It was found that there was significant correlation between the brain region activation and the musical impact regions predicted. Control analysis with loudness measurements was also conducted. Further, participants indicated higher impact during phrases predicted to be chill-eliciting by MIIR.

Critique:
The hypothesis and argument is convincing because music employs many tools like key changes, melodic peaks, changes in loudness, tempo, etc to elicit reactions in listeners. Thus, the software relies on these musical attributes and identifies the impact regions. Results of the study also show a strong correlation between the predictions and the brain scans ($p < 0.01$).

Key Insight:
The study uses two different methodologies to test their hypothesis and also make sure to factor out loudness in the brain study. This makes sure that the brain activation levels we see are not only because of the loudness of the music.
One limitation of the study is that the fMRI imaging was only performed on non-musically trained individuals which could impact generalisability.

b. **Artist2Risk: Predicting Depression Risk based on Artist Preferences** (Yash Goyal, Vinoo Alluri)

Summary:
Building on previous research that analyses the acoustic features and social tags of music listened to by individuals at-risk for depression (via the online music streaming platform Last.fm), this study aims to identify artist diversity and associate it to listeners' risk for depression

and explore the potential to predict the depression risk using artist-based preferences on Last.fm.

Surana et al.(2020) used 541 participants' Last.fm data over six months and their Kessler's Psychological Distress (K10) scores. To factor in the tendency of an individual to listen to diverse artists, an exploration score per participant was computed as the ratio of the number of unique artists to the total play-count in that period. Mann-Whitney-U Tests, followed by permutation tests were used to capture differences between the No-Risk (K10 $\leq$ 19) and At-Risk (K10 > 29) groups. An SVM model was trained to classify the participants into one of the two groups using a feature vector extracted from this data.

A significant difference ($p < 0.001$) was found between the No-Risk and At-Risk group. It is found that the No-Risk group tends to explore more diverse artists. The classification accuracy with 10-fold cross-validation was found to be 73.11% which is considerably higher than chance level (50%).

Critique:
The study finds that people at high risk for depression have a tendency to not explore new artists while exploring music. This argument is convincing because it is also backed by previous research on depression that states depression risk is negatively correlated with diverse preferences. The classification accuracy of the model is also an indicator that artist preferences could be correlated with depression risk.

Key Insights:
The study uses statistical methods to quantify the exploration scores of the listeners. This helps build a very reliable and accurate model that predicts the depression risk of individuals.
One limitation of this study could be the effect of confounding variables like social and cultural influences on the artists' that one chooses to listen to.

c. **Acoustic Vowel Space in the Speech-to-Song Illusion** (Anna Hiemstra, Makiko Sadakata)

Summary:
Due to the acoustical differences between speech and song, like rhythm and pitch, it is easy for us to distinguish between them. However, in the Speech-to-Song illusion, a spoken phrase starts to sound like song, indicating that the difference between speech and song is purely perceptual. Previous research looked into rhythmic and melodic aspects of transforming and non-transforming sentences. This study aimed to study the relation between the timbral quality of the sentences, more specifically - the vowels, and how the audio is perceived.

Formants are frequency resonances produced by the vocal tract and they are used to characterise vowels. The vowel space area is the 2D plane formed by plotting these frequencies. Analysing this space tells us about the acoustic characteristics of the vowels in different conditions. The size of the acoustic vowel space area bounded by formants of /i/,/æ/, and /ɔ/ in a data set of English sentences that either strongly transform to song or remain perceptually stable as speech is analysed. The Euclidean distances of individual vowel tokens in song and speech are compared to the circumcenter of the combined stimuli to quantify

if there is a relationship between the vowel space area and speech or song perception.

The mean Euclidean distances between individual vowels and the overall circumcenter were found to be significantly larger in speech than in song, suggesting that the vowel space area of speech is larger than that of song. Thus, vowel formants with more extreme frequency values may be perceptually associated with speech.

Critique:
The argument is convincing because it is backed by a lot of previous research that suggests that the vowel space area for speech and song are both different and in the speech-to-song illusion, the vowel space closely resembles the vowel space associated with song. Extreme formant values can consequently activate speech circuitry, in that way resisting the perceptual transformation to song.

Key Insights:
The main strength of the study is that it has a large impact on our understanding of speech and music because it identifies and quantifies a new feature that is used to distinguish between the two.
The study was only conducted with '/i/,/æ/, and /ɔ/' vowels so the results cannot fully represent the whole vowel space.

d. **The Active Minds Music Ensemble: Learning music online and in person** (Jennifer MacRitchie, Anthony Chmiel, John R. Taylor, Madeleine Radnan, Catherine J. Stevens, Roger T. Dean)

Summary:
Learning any musical activity is linked to a large range of benefits, especially considering all the research regarding the transfer learning

effects of music. Quantitative evidence, however, shows low to moderate general cognitive benefits for older adults taking part in short-term (between 2-4 months) music instrument training programs, suggesting that a longer duration of training may be necessary to see a transfer of skills.

Thus, *The Active Minds Music Ensemble* aims to answer this question by conducting a longitudinal research study on how older adults' can benefit from music learning. The project aims to determine the extent of transfer to domain-general cognitive skills, their retention of these learned skills and assess the cognitive load involved in the learning process.

The healthy older adult participants receive 60 minute music lessons for 12 months, once every two weeks. All participants learn on a digital keyboard and also the iPad app, Thumbjam, spending 6 months on each instrument. They are all taught by ear without using any notation and activities involve performing single-line melodies and improvisation.

After the lessons, various tests are conducted to assess cognitive skills (digit span test, alternate uses task), motor function (Finger Tapping Test and Trail Making Test), musical aptitude (a perceptual pitch direction task, and the Goldsmith's Musical Sophistication Index, including melodic discrimination and beat perception tasks), and aspects of wellbeing (General Self-Efficacy, CASP-12 and Basic Psychological Needs Scale). Qualitative interviews will also be conducted to record the participants' own perception of learning and musical ability.

Since the study is still ongoing, there are no tangible results and conclusions drawn as of yet. A preliminary sub-sample of older adults' (n=52) were asked about their experiences, and most participants are surprised at their ability to learn a musical instrument, especially online.

Critique:
Since there are no results yet, I cannot fully agree or disagree with any reasoning. However, I do think that a longer duration of training will result in a better transfer learning of musical ability in older adults', as already qualitatively indicated by the small subset of participants.

Key Insights:
A longitudinal study lasting over a year, will result in conclusive and accurate results about the effect of short-term music programs on older adults'.
The only limitation of this study could be the effect of confounding variables like practice time, lifestyle and social interactions that could possibly affect the results of this study.

1.2)

1. **Acoustic Vowel Space in the Speech-to-Song Illusion** (Anna Hiemstra, Makiko Sadakata)

Local Adaptation

Due to the multilingual population in IIIT Hyderabad, we would need to account for how the Speech-to-Song illusion would work in a language other than English - say Telugu. Including native languages can help

explore if the illusion is more pronounced in a familiar or unfamiliar language. Existing native language datasets can be used to form the transforming and non-transforming sentences.

Thus, to adapt the study to a IIIT context - we can use native languages and include multilingual people in the study. We will also need to account for the different vowel contexts in Indian languages and how they would affect the Speech-to-Song illusion.

Outcome Variation

The participant demographic in IIIT will mostly consist of educated students, with technology access. Furthermore, almost all of them will be multilingua. Due to this, there could be a difference in the vowel space area in native languages as compared to English, thus affecting the illusion.

While there has been research on occurrence of the illusion in Bangla speakers, vowels might differ across Indian languages. Further, Indian languages are syllabic while English is alphabetic - leading to differences in how vowels and words are pronounced. This could cause a difference in how the illusion is perceived.

Since Indian languages are syllabic there could be less stress on the vowels during pronunciation, leading to lower frequency for vowels, and as found by the paper, being more likely to be perceived as song. Thus, the outcome may show more susceptibility to the illusion due to multilingualism.

Further Inquiry

Hypothesis question: Is the Speech-to-Song illusion more perceived in familiar or unfamiliar languages?

We can get participants in IIIT who are more familiar in their native language than in English and then see how the illusion is perceived by them. This will give us insights into the cognitive process of language familiarity.

2. **The Effects of Interleaved and Blocked Practice on Musical Style Recognition** (Elizabeth Kinghorn, Rebecca Lagace-Cusiac, Ozgen Demirkaplan, Jessica A. Grahn, Jonathan de Souza, Christine Carter)

Local Adaptation

If this study was to be conducted in IIIT Hyderabad, we would have differences in the type of music that most people would be exposed to. The musically trained participants could be trained only in Indian classical music, while some may be trained in Western music and this distinction could make a difference.

Further, the stimulus audio presented could itself be changed to include the cultural context of the region's music. In addition to the Western classical music, we can also add Indian music to the stimulus and compare any differences in the results.

Outcome Variation

Since both groups of participants (musically trained and no prior training) from IIIT would be familiar with Indian music, they might have an easier time recognising the Indian composers as compared to the Western composers. If we were comparing only the differences observed in Western music, the results may not differ as much - with interleaved learning proving the most beneficial for musicians. The only difference, if any, could be if the musicians were only trained in classical Indian music and thus, the learning curve for both musicians and non-musicians would be the same for the Western music stimulus.

Further Inquiry

Hypothesis question: Does familiarity with the music genre (Indian classical and Western classical) affect the ability to recognize the composer in addition to the type of practice (blocked vs interleaved)?

Since participants from IIIT will have varying degrees of familiarity to both Indian and Western music, we can test whether this is a factor to consider. The results will tell us more about the relationship between prior familiarity to concepts and the effectiveness of the different practice schedules.

Use of AI tools
Used AI tools (ChatGPT) to simplify abstracts that I could not understand easily and to also formulate my thoughts about certain topics.