CHILLS AND BETTING BILLS – AUDIO FEATURES, LISTENER EMOTIONS AND THE SUCCESS OF ESC SONGS

Helene Lindenbauer[1], Christoph Reuter[1], Sarah Ambros[1]

[1] *SInES, Musicological Department, University of Vienna, Vienna, Austria*

This contribution has been accepted based on the submitted abstract. The submitted manuscript was not reviewed and is reproduced here without edits or corrections by the conference board. Neither the EAA nor the AAA take responsibility for its contents.

Abstract

Each year, the Eurovision Song Contest (ESC) generates speculation about the winning song. Previous research about Eurovision Song Contest Songs has mainly examined the musical features of past ESC participants, while the listeners' emotions and chill experiences have not yet been combined with it. This study investigated whether acoustic and emotional differences can be measured between highest- and lowest-rated ESC songs based on pre-final betting odds (April 2026). Twenty-seven participants listened to three lowest- and three highest-rated songs, while skin conductance was measured via Mindfield eSense sensors and valence, arousal (Circumplex Model) and experienced chills were reported. After each song, participants were asked to rate the song liking and how often they have heard it before. Further, they filled out a personality questionnaire (SEPPO). Audio features were extracted using a.o. MIRToolbox and PADMEA. There were significantly more experienced chills for the three higher-rated songs ($t_{(3,first\_vs.\_3,last)}=3.264$, $p<0.05$). Differences (t-Tests) between personality traits, song liking, valence, arousal, and musical audio features between the three lowest and three highest-rated songs were noticed. Based on the acoustic audio feature analysis the valence also depends on the songs Roughness (Vassilakis, $t_{(3,first\_vs.\_3,last)}=-7.528$, $p<0.01$) and Brightness ($t_{(3,first\_vs.\_3,last)}=11.935$, $p<0.01$).

Keywords: *audio signal analysis, chill experience, emotions, listening task*

1. Introduction

The Eurovision Song Contest (ESC) is one of the biggest music events in the world, gathering millions of viewers and listeners from all over the globe. Each participating country decides on one song and artist to be performed at the ESC. The winning song is determined in three steps: (1) competition in a semi-final, which decides participation in (2) the final, where all qualified songs are performed, and (3) a combination of jury and public votes, which ultimately determines the winning song. Every year, betting odds attempt to predict which song will win. In addition, researchers have tried in recent years to identify factors that influence which song wins.

A multimodal approach to Hit Song Science (HSS) applied to the ESC showed that the best prediction of a winning song requires an analysis of the songs' intrinsic characteristics, public appeal, and Eurovision-specific data, while noting that no single factor alone is capable of predicting outcomes [1]. There is also evidence that circumstances unrelated to the song itself have an influence, such as running order [2], cultural dynamics [3], and the voting system [4]. It has also been shown that songs are adapted and manipulated to fit the criteria of past winning songs, in an attempt to increase their chances of winning the contest [5].

2. Research Questions

While existing studies focus on musical features and cultural and economic influences on past ESC songs, listeners' emotions and their experience while listening to these songs remain largely unexplored. This study aims to determine whether there are differences in listeners' emotional responses when listening to the highest- and lowest rated ESC songs. Several approaches were used to examine these differences. First, listeners tracked their emotions throughout each song using an interactive circumplex model [6]. Second, they reported experienced chills, since these represent a psychophysiological reaction associated with emotional states and are linked to pleasure, as well as to whether something is perceived as aesthetic [7]. Third, participants completed a short personality questionnaire (SEPPO) [8], as research indicates differences in preference and physiological responses depending on personality traits [9]. Musical audio

Corresponding author: *helene.lindenbauer@univie.ac.at.*

features were also extracted in order to link them to the emotional experiences reported.

3. Participants

27 participants took part in this study. The participants were recruited mostly from the University of Vienna via word of mouth. 17 participants identified as female, 10 as male. Their ages ranged from 19 to 57 years, with a mean age of 28 years. The study took place on site, with a maximum of four participants taking part at the same time. Participants were briefed on the task before the listening session began. Since knowledge of the study's aim could have influenced their responses, this information was only disclosed after the experiment had concluded.

4. Methods

4.1 Music stimuli and participants scoring of the songs

The stimuli were chosen based on pre-final betting odds from April 2026, which indicated which of the submitted Eurovision Song Contest 2026 entries were considered most and least likely to win. The three highest-rated songs (from highest to third highest) were: "Liekinheitin" by Linda Lampenius & Pete Parkkonen (Finland), "Regarde!" by Monroe (France) and "Før vi går hjem" by Søren Torpegaard Lund (Denmark). The three lowest-rated songs were "Nova Zora" by Tamara Živković (Montenegro), "To epic to be true" by Vanilla Ninja (Estonia) and "Tanzschein" by Cosmó (Austria). Live recordings were chosen, since the Eurovision Song Contest is also performed live. Amplitudes were normalized so that the volume levels were approximately consistent across all participants and pieces of music. Participants listened to the songs via electroacoustic headphones, presented in randomized order. After each song, participants answered one question on how much they liked the song, on a scale from 1 (not at all) to 4 (very much), and a second question on how often they had previously heard the song, ranging from 1 (never) to 4 (many times). After listening to all songs, participants completed a personality questionnaire and answered two open questions regarding their general opinion of the Eurovision Song Contest and the music performed there.

4.2 Measurement of chills, valance, arousal and skin conductance

While listening to each of the songs, the participants had two main tasks:

- They had to press and hold the left mouse button for the duration of the chill experience, releasing it once the sensation had subsided.
- They had to report their valence and arousal by moving the mouse cursor over a circumplex model [6].

For these tasks, a web application from the SInES Tools suite was used [10]. In addition to participants' self-reports, skin conductance was measured via a Mindfield eSense sensor, which was likewise recorded through the same web application [10].

4.3 Evaluation of musical audio features

Musical audio features had been extracted using the MIRToolbox [11], and PADMEA [12], to assess influences of these audio features on song liking, chills, skin conductance and emotions. The raw data for each participant and each song, recorded with the SInES web application [10], was exported to Microsoft Excel, where mean values were calculated for each parameter. T-tests were conducted to assess differences between the parameters and the higher- and lower-rated songs, and Spearman's correlations were conducted to assess correlations between the different parameters. For both t-tests and Spearman's correlations, Shapiro-Wilk tests were additionally computed.

5. Results

5.1 Sample description

Information regarding participants' liking of the six different songs, whether they had heard the song before, and the mean values of the personality questionnaire (SEPPO) [8] is shown in table 1. Decisions on high or low scores for each variable were made by dividing the data into two groups based on whether they were above or below the mean value. Very few of the participants had heard one of the songs before, which indicates a higher likelihood of experiencing chills, as research points to a decrease in chills with increased listening time [13]. There is no significant effect of high or low song ranking on song liking ($t_{(3,first\_vs.\_3,last)}=2.162$, $p<0.1$), but the two highest-rated songs (Finland (mean=2.39) and France (mean=2.82)) have the highest scores in song liking.

Table 1 Mean and SD scores from the participants' questionnaires, indicating whether they liked and knew the song played, along with the personality questionnaire results.

Variable	Mean	SD	Range
Song liking	2.12	0.44	1.53-2.82
Songs heard before	0.15	0.29	0-0.071
Extraversion	18.68	2.97	13-25
Neuroticism	16.29	3.11	10-24
Psychoticism	13.74	2.63	9-18
Openness	24.04	2.58	19-28

5.2 Differences between highest and lowest rated songs and gathered data

5.2.1 Subjective chill experiences

A higher amount of experienced chills was related to the higher-rated songs (mean=0.031, SD=0.015), while fewer chills were perceived for the lower-rated songs (mean=0.005, SD=0.004) ($t_{(3,first\_vs.\_3,last)}=3.264$, $p<0.05$). Figure 1 shows a comparison of the perceived chills for the songs "Regarde!", which elicited the

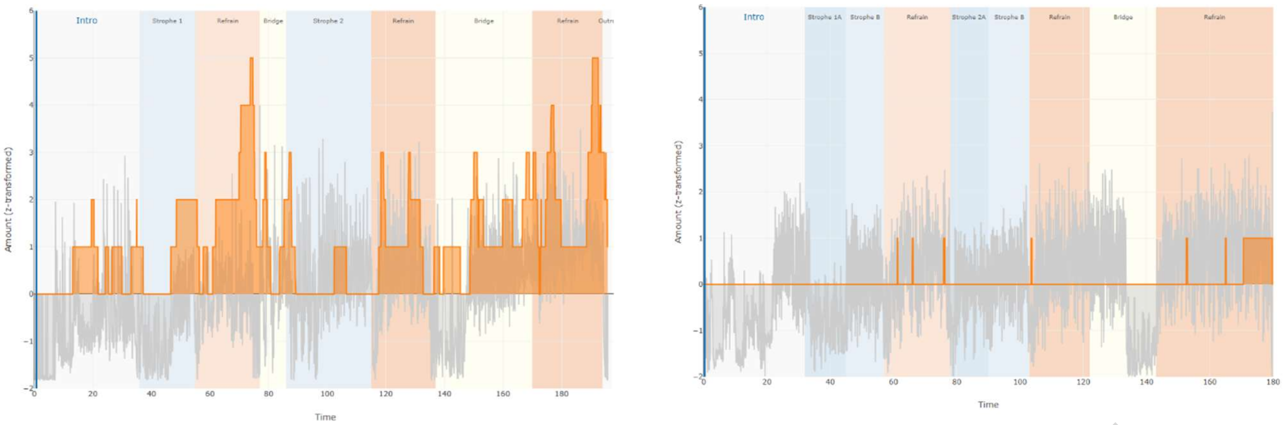

Figure 1 Perceived chills for the songs from France (left) and Estonia (right), shown as bars, Amplitude is visible in the background. X-axis shows the time, y-axis the number of perceived chills from all participants throughout the song. Vertical sections are dividing the songs in Intro, Verse, Bridge, Chorus and Outro.

highest number of chills, and "Too Epic to Be True", which elicited the lowest number of chills.

Research indicates a higher likelihood of experiencing chills at transitions from chorus to verse, between bridge and chorus, or when something new happens suddenly and unexpectedly, such as a rapid change in loudness or the introduction of new instruments [13], [14], [15]. These known factors are also important indicators within the chills reported in this study. As shown in Figure 1, experienced chills mostly occur at unexpected, high-contrast events. One such instance was reported by five participants at the end of the first chorus, where a sudden pause follows the opera-like singing and the dense instrumentation preceding it. Comparing the lower-rated songs to the higher-rated songs, the former show fewer dynamic, tempo-related or other sudden and unexpected changes.

5.2.2 Emotional Differences and skin conductance

No significant differences in emotional values, assessed through the circumplex model, were found between the highest- and lowest-rated songs. The same applies to skin conductance responses between the higher and lower rated songs, despite an assumed higher response for the higher-rated songs, since higher skin conductance is associated with the experience of chills [9].

5.2.3 Audio features

A significant difference in roughness (Vassilakis) was found between lower rated (mean=4.77, SD=0.36) and higher rated songs (mean=7.92, SD=0.82) ($t_{(3,\text{first\_vs.\_3,last})}=-7.528$, $p<0.01$). Roughness has previously been associated with dissonant and unpleasant sounds [16], suggesting that songs with higher roughness scores are more likely to be perceived as unpleasant, which may have influenced participants' liking of the songs. Furthermore, a significant difference was found for tempogram stability between higher rated (mean=0.08, SD=0.04) and lower rated songs (mean=0.31, SD=0.09) ($t_{(3,\text{first\_vs.\_3,last})}=-4.85$, $p<0.01$). As discussed in chapter 5.2.1, more unexpected and sudden events are associated with a higher likelihood of experiencing chills. The higher tempogram stability values found in the lower-rated songs therefore could reflect a perception of fewer tempo-related changes. A significant difference was also found for brightness between higher rated (mean=0.51, SD=0.01) and lower rated songs (mean=0.46, SD=0.01) ($t_{(3,\text{first\_vs.\_3,last})}=11.935$, $p<0.01$). Higher Brightness values have been associated with a higher probability of experiencing chills [17].

5.3 Correlations between the liking of the songs, emotions, chills and personality

In addition to the differences between the lowest- and highest-rated songs across the measured parameters, correlations between those parameters were computed. The data in Table 2 were derived from the personality questionnaire (SEPPO) [8] and from mouse movements over the circumplex model within the web application [10]. It shows the correlations between experienced chills, different emotions, overall song liking, and high or low scores on the personality questionnaire (SEPPO) [8].

Table 2 Correlations between different parameters. First row: data from high/low personality scores (E=Extraversion, N=Neuroticism, P=Psychoticism, O=Openness) combined with Chills (C). First Column: participants' liking of the song and the different emotions computed from the Circumplex Model (CM)

	High E + C	Low N + C	High P + C	High O + C
Song Liking	$r_s=0.94$ $p<0.01$	$r_s=0.82$ $p=0.05$	$r_s=0.87$ $p<0.05$	
Cm Valence	$r_s=0.89$ $p<0.05$		$r_s=0.89$ $p<0.05$	
Cm Pleased	$r_s=0.94$ $p<0.01$			
Cm Peaceful	$r_s=0.94$ $p<0.01$		$r_s=0.89$ $p<0.05$	
Cm Relaxed	$r_s=0.89$ $p<0.05$			
Cm Bored			$r_s=-0.83$ $p<0.05$	

Cm Angry	r_s=-0.83 $p<0.05$	
Cm Nervous		r_s=0.89 $p<0.05$
Cm Excited		r_s=0.83 $p<0.05$

6. Discussion

Results indicate significant differences between higher- and lower-rated songs regarding chill experiences and audio features. The higher-rated songs elicited a significantly greater number of experienced chills compared to the lower-rated songs. Musical features associated with chills, as explored in previous research [13], [14], [15], were also identified as important factors in this study. Higher-rated songs featured a greater number of chills, more unexpected and sudden changes, lower scores in roughness and tempogram stability and higher scores in brightness. Considering research indicating that no single feature can exclusively predict the outcome of the ESC [1], this study identified significant features of higher-rated ESC songs that could potentially contribute to their success. Since the study took place before the ESC semi-final and final, and the classification into higher- and lower-rated songs was based on pre-final betting odds, it was of interest how the songs would ultimately perform in the contest and whether this classification would prove accurate. Two of the lower-rated songs (Montenegro and Estonia) did not advance past the semi-final and therefore had no opportunity to perform in the final. The third lower-rated song (Austria) did reach the final without having to perform in the semi-final, since the host country automatically qualifies for the final. Nevertheless, this song received the second-to-last place in the final. None of the higher-rated songs won the ESC, but they achieved 6th (Finland), 7th (Denmark), and 11th (France) place, placing them considerably ahead of the lower-rated songs.

7. Conclusion

This study aimed to identify differences between higher- and lower-rated songs of the ESC 2026, based on pre-final betting odds. Differences were identified and could potentially contribute to the final outcomes. The study shows that considering listeners' perception, their psychophysiological reactions and audio features when investigating the winning chances of future ESC songs could contribute positively to this field of research. For further research, it would be interesting to include a greater number of participating ESC songs to enhance the results. For interactive results visit https://muwiserver.univie.ac.at/esc2026/.

References

[1] K. Adamska and J. Reiss, “Predicting Eurovision Song Contest Results A Hit Song Science Approach,“ in Transactions of the International Society for Music Information Retrieval, vol. 8, no. 1, May, pp. 93–107, 2025.

[2] J. L. Arroyo-Barrigüete, L. Fernández and A. Obregón, “Eurovision in the boardroom: What does voting order tell us about decision-making?,” in The Eurovision Song Contest as a Cultural Phenomenon, A. Dublin, D. Vuletic and A. Obregón, Eds. London: Routledge, 2022, pp.144-159.

[3] D. García and D. Tanase, “Measuring Cultural Dynamics Through the Eurovision Song Contest,” in arXiv, Art. no. 1301.2995v2, Nov. 2024. [Online serial]. Available: https://doi.org/10.48550/arXiv.1301.2995. [Accessed: July 4, 2026]

[4] V. Ginsburgh and J. D. Moreno-Ternero, “The Eurovision Song Contest: voting rules, biases and rationality,” in Journal of Cultural Economics, vol. 47, no. 2, June, pp. 247-277, 2023.

[5] L. A. Nunes Amaral, A. Capozzi and D. Helbig, “Breaking the Code. Multi-level Learning in the Eurovision Song Contest,” in *arXiv*, Art. no. 2505.11555, May 2025. [Online serial]. Available: https://doi.org/10.48550/arXiv.2505.11555. [Accessed: July 4, 2026].

[6] J. A. Russell, “A circumplex model of affect,” in *Journal of Personality and Social Psychology*, vol. 39, no. 6, Dec., pp. 1161–1178, 1980.

[7] R. de Fleurian and M. T. Pearce, “Chills in Music: A Systematic Review,” in *Psychological Bulletin*, vol. 147, no. 9, Sept. pp. 890-920, 2021.

[8] R. von Georgi and J. Herr, “SEPPO: Short Eysenck Personality Profiler with NEO-PI-R-Openness,”. Berlin Institute of Biomusicology and Empirical Research (BIBER), Berlin, 2020.

[9] K. Starcke, R. von Georgi, T. M. Tiihonen, K.-F. Laczika and C. Reuter, ”Don’t drink and chill. Effects of alcohol on subjective and physiological reactions during music listening and their relationships with personality and listening habits,” in International Journal of Psychophysiology, vol. 142, Aug., pp. 25–32, 2019.

[10] C. Reuter, SInES Tools: Emotion and Audio Features Analysis, 2021. [Web Application]. Available: https://sinestools.univie.ac.at/va_tracker.htm [Accessed: July 3, 2026].

[11] O. Lartillot and P. Toiviainen, “A Matlab toolbox for musical feature extraction from audio,” in Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, Sep. 10–15, 2007.

[12] I. Czedik-Eysenberg, “Semantische Modellierung wahrnehmungspsychologischer Musikdimensionen auf Basis von akustischen

Signaleigenschaften,“ Dissertation, University of Vienna, Vienna, Austria, 2021.

[13] O. Grewe, F. Nagel, R. Kopiez and E. Altenmüller, „Listening To Music As A Re-Creative Process: Physiological, Psychological, And Psychoacoustical Correlates Of Chills And Strong Emotions,” in Music Perception: An Interdisciplinary Journal, vol. 24, no. 3, Feb., pp. 297-314, 2007.

[14] S. Bannister, “A Vigilance Explanation of Musical Chills? Effects of Loudness and Brightness Manipulations,” in Music and Science, vol. 3, pp. 1-17, 2020.

[15] J. Panksepp, “The emotional Sources of ‘Chills’ induced by music,” in Music Perception, vol. 13, no. 2, Dec., pp. 171-207, 1995.

[16] E. Terhardt, “The Concept of Musical Consonance: A Link Between Music and Psychoacoustics,” in Music Perception, vol. 1, no. 3, April, pp.276-295, 1984.

[17] E. J. Beier, P. Janata, J. C. Hulbert, F. Ferreira, „Do You Chill When Chill? A Cross-Cultural Study of Strong Emotional Responses to Music,” in Psychology of Aesthetics, Creativity, and the Arts, vol. 16, no. 1, Feb., pp. 74-96, 2022.