Examensarbete inom medieteknik

Grundnivå, 15 hp

Promptingstrategier och emotionell precision

En kritisk undersökning av mänsklig intention vid användning av AI-genererad musik

Max Kjellgren

Promptingstrategier och emotionell precision: En kritisk undersökning av mänsklig intention vid användning av AI-genererad musik

Max Kjellgren
EECS School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology,
maxkjel@kth.se

This study investigates the emotional precision of music generated by the AI model Suno v4.5 by comparing two strategies: *semantic emotional prompting* and ***s**tructural parametric prompting*. Drawing on Russell's *Circumplex Model of Affect*, the ability to convey the emotions of calm, tension, and sorrow within an audiovisual context was evaluated through an experimental design involving 31 participants. The results indicate that the strategies are generally equivalent in steering emotional intention, although the semantic strategy proved statistically significantly more effective at conveying sorrow. The highest emotional precision was recorded for the category of tension, followed by calm, while sorrow exhibited the study's lowest precision values. The study discusses a potential tendency of the AI model to generate music with a positive valence bias, which may complicate the expression of dark or subdued emotions. Furthermore, the significant impact of the visual material on how the music is interpreted is highlighted, suggesting that AI-generated music is most effective when there is a clear interplay with the visual components. The work problematizes the shifting creative process as manual craftsmanship is replaced by algorithmic suggestions. The user's role is seen to shift from that of a traditional composer to more closely resemble a requester, raising existential questions regarding the value of artistic creation and the preservation of a personal touch within the creative process.

CCS CONCEPTS

- **Applied computing ~ Sound and music computing**
- **Computing methodologies ~ Artificial intelligence**
- **Human-centered computing ~ Human computer interaction (HCI)**

Additional Keywords and Phrases: Generative AI, Music generation, Emotional precision, Suno v4.5, Prompting strategies, Human-AI interaction, Audiovisual congruence.

Sammanfattning

Denna studie undersöker den emotionella precisionen i musik genererad av AI-modellen Suno v4.5 genom att jämföra två strategier: *semantisk emotionell prompting* och *strukturell parametrisk prompting*. Med utgångspunkt i Russells *Circumplex Model of Affect* utvärderades förmågan att förmedla känslorna lugn, spänning och sorg i en audiovisuell kontext genom en experimentell design med 31 deltagare. Resultaten visar att strategierna överlag är likvärdiga för att styra emotionell intention, men att den semantiska strategin är signifikant mer effektiv att förmedla sorg. Högst emotionell precision uppmättes för kategorin spänning, följt av lugn, medan kategorin sorg uppvisade studiens lägsta precisionsvärden. Studien diskuterar en möjlig tendens hos AI-modellen att skapa musik med positiv valens, vilket kan försvåra uttryck av mörka eller dämpade känslor. Vidare understryks bildmaterialets stora inverkan på hur musiken tolkas, där resultaten tyder på att den AI-genererade musiken fungerar som mest effektivt när det finns ett tydligt samspel med det visuella materialet. Arbetet problematiserar hur den kreativa processen förändras när den hantverksmässiga

ansträngningen ersätts av algoritmiska förslag. Användarens roll skiftar från traditionell kompositör till att mer likna en beställare, vilket exempelvis väcker existentiella frågor om det konstnärliga skapandets värde och den personliga prägelns bevarande i skapandeprocessen.

1. Introduktion

Den snabba framväxten av generativa AI-system har på kort tid fått stor påverkan på hur kreativt innehåll skapas. Inom branscher som filmproduktion, marknadsföring, musikproduktion och spelutveckling används olika typer av AI-genererat material i allt högre grad. Tekniken kan utnyttjas både i produktionens huvudfokus och för att hantera sekundära men nödvändiga aspekter såsom musik. Exempelvis kan filmproducenter och spelutvecklare generera skräddarsydd musik för att effektivisera de delar av produktionen som ligger utanför deras eget expertområde. Detta minskar produktionskostnaderna och kan tillåta skapare att självständigt färdigställa kompletta audiovisuella projekt utan att vara beroende av externa kompositörer. Inom just musikproduktionen har nya möjligheter öppnats upp för användare utan musikteoretisk bakgrund att skapa kompletta kompositioner. Genom verktyg som Suno[1] och Udio[2] kan numera musik genereras direkt utifrån språkliga instruktioner. Samtidigt har musiken en helt avgörande roll för hur rörlig bild tolkas, då den har förmågan att forma förståelsen för en specifik scen, skapa spänning eller förstärka avsedda emotionella uttryck [6]. För en audiovisuell skapare är det därför av yttersta vikt att kunna styra dessa uttryck. I en traditionell skapandeprocess sker denna styrning genom att en musiker eller kompositör själv formar musiken genom direkt konstnärligt hantverk, medan exempelvis en filmregissör eller spelutvecklare kan styra processen genom att kommunicera sin vision till en extern kompositör. När musikskapandet istället sker genom interaktion med en text-till-musik-modell förändras denna process i grunden, oavsett användarens bakgrund. Eftersom interaktionen bygger på textbaserade instruktioner (prompts) sker användarens kontroll över det musikaliska resultatet på ett indirekt sätt. Användarens roll ändras från att kontrollera detaljer eller samarbeta med en extern kompositör, till att själv styra ett generativt system. Precisionen i det emotionella uttrycket blir därmed till stor del beroende av hur instruktionerna till systemet formuleras.

Interaktionen med text-till-musik-modeller kan utformas utifrån olika strategier beroende på hur användaren väljer att formulera sina språkliga instruktioner. En metod är att beskriva den avsedda emotionen med hjälp av enkla och abstrakta begrepp som ”lugn”, ”spänning” eller ”hotfull”. Detta tillvägagångssätt, som i denna studie benämns som *semantisk emotionell prompting*, gör verktygen tillgängliga även för användare utan musikteoretisk bakgrund. Detta beror på att användaren kan förlita sig på ett enkelt språkligt gränssnitt för att förmedla en specifik känslostämning, utan att behöva besitta teoretisk kunskap om specifika musikaliska parametrar och deras betydelse.

En alternativ kommunikationsstrategi är istället att specificera teoretiska parametrar som exempelvis instrumentation, tempo eller dynamik. Denna metod, som i studien benämns som *strukturell parametrisk prompting*, kräver en högre grad av medvetenhet kring hur musikens uppbyggnad påverkar det emotionella uttrycket. Valet mellan att använda abstrakta beskrivningar eller teoretiskt specificerade instruktioner innebär därför ofta en avvägning mellan kontroll och användarvänlighet [14].

[1] https://suno.com/
[2] https://www.udio.com/

Detta behov av kontroll är centralt inom fältet människa-datorinteraktion, där man betonar vikten av att interaktiva system är begripliga och ger användaren en känsla av förutsägbarhet [15]. I traditionella musikredigeringsprogram kan användaren manipulera enskilda ljudparametrar med direkt feedback, medan interaktion med generativ AI istället innebär en mer indirekt kontroll då relationen mellan användarens intention och systemets output inte längre är lika transparent och momentant. Små variationer i en formulering kan resultera i stora skillnader i slutresultatet, och det kan vara otydligt hur systemet översätter de semantiska begreppen till konkreta musikaliska egenskaper.

Bristen på transparens och kontroll är bara en av få konsekvenser av att använda AI-modeller för att skapa musik. Det finns även djupare etiska och existentiella frågor kring vilken roll AI-musik kan få och vilka konsekvenser det kan få för samhället och kreativa processer i stort. Exempelvis finns det en oro för att professionella yrkesroller som musiker, producenter och kompositörer försvinner när musikproduktionen effektiviseras genom algoritmer. Även den konstnärliga kvaliteten ifrågasätts när musik genereras fram snarare än arbetas fram hantverksmässigt. Holzapfel et al. [10] tar upp problematik kring upphovsrätt, som exempelvis att AI-modellerna tränas på befintlig musik utan att ta hänsyn till ursprungsskaparnas intressen. Denna konflikt har i praktiken lett till omfattande rättsliga processer, där stora delar av musikindustrin har stämt just företag bakom modeller som Suno och Udio för storskaligt upphovsrättsintrång gällande träningsdata [4]. Vidare lyfts även frågan om kulturell appropriering i form av risken att AI-genererad musik riskerar att trivialisera och skada musiktraditioner [10]. Det finns också diskussioner kring hur AI kan användas för att faktiskt stödja den kreativa processen. Exempelvis menar Anantrasirichai och Bull [1] att den största nyttan med AI inom de kreativa branscherna uppstår när tekniken är människocentrerad. De menar att fokus bör ligga på att utforma system som förstärker snarare än ersätter mänsklig kreativitet.

1.1 Syfte och forskningsfråga

Syftet med denna studie är därför att kritiskt undersöka i vilken utsträckning ett avsett emotionellt uttryck kan kommuniceras genom generativ AI-musik. Genom att fokusera på användarens kontroll i en audiovisuell kontext undersöks hur väl olika promptingstrategier kan översätta en specifik emotionell vision till ett konkret resultat.

Utöver att analysera musikens emotionella precision syftar arbetet till att analysera och problematisera den roll människan tilldelas i skapandeprocessen, men även att bidra till diskussionen kring de etiska och existentiella konsekvenser som kan uppstå när mänsklig intention och konstnärligt uttryck automatiseras.

Utifrån syftet har följande forskningsfråga formulerats:

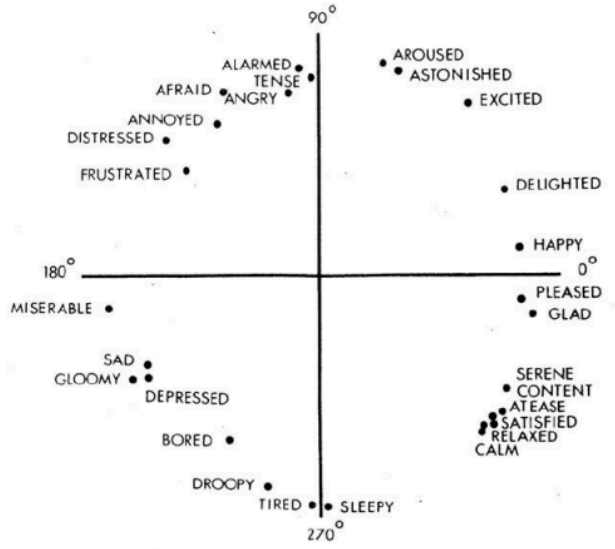
Vilken av promptingstrategierna semantisk emotionell prompting eller strukturell parametrisk prompting ger högst emotionell precision i en audiovisuell kontext?

2. Bakgrund

I detta kapitel presenteras den teoretiska ram och tidigare forskning som ligger till grund för studiens undersökning. Inledningsvis behandlas teorier kring musik och emotioner, med fokus på hur känslor kan kategoriseras och mätas genom etablerade psykologiska modeller. Därefter beskrivs hur musik används i audiovisuella sammanhang för att framkalla känslor och styra narrativ. Slutligen diskuteras framväxten av AI inom musikproduktion och flera tidigare studiers resultat kring dess emotionella precision presenteras.

2.1 Emotionell perception av musik

Musikens förmåga att kommunicera och uttrycka emotionella karaktärer är ett välstuderat område. Inom musikpsykologin skiljer man mellan känslor som faktiskt väcks hos lyssnaren och de känslor som lyssnaren identifierar i musiken. Denna studie avgränsas helt till deltagarnas förmåga till emotionell perception och igenkänning av den avsedda känslan i det audiovisuella materialet . Enligt Juslin och Västfjäll [12] uppstår dessa känslor genom att musikaliska strukturer interagerar med både kognitiva och fysiologiska system hos individen. För att beskriva emotionella responser kan man dels använda sig av kategoriska modeller som till exempel glädje, sorg eller rädsla. Men man kan även använda sig av så kallade dimensionella modeller, där en av de mest etablerade modellerna är Valens–arousal-modellen (se Figur 1). I denna modell representeras valens behaglighet på den horisontella axeln (från negativ/obehaglig till positivt/behaglig), medan arousal (energinivå/intensitet) representeras på den vertikala axeln (från låg till hög aktivering).



Figur 1: Valens–arousal-modellen (Russells cirkulära modell för affekt, [16])

Forskning visar även att specifika musikaliska parametrar korrelerar med hur emotionellt uttryck uppfattas. Gabrielsson och Lindström [9] sammanfattar exempelvis hur tempo, tonart, dynamik och instrumentation påverkar upplevd valens och arousal. Exempelvis associeras snabbt tempo och hög dynamik med hög arousal och durtonarter ofta med högre valens än molltonarter. Det bör dock noteras att även om studier indikerar att vissa grundläggande mönster i tonalitet kan kännas igen tvärkulturellt [8], betraktas kopplingen mellan dur/moll och specifika känslor ofta som en västerländsk musikalisk konvention som förstärks genom kulturell inlärning [9].

För att kunna styra musikaliskt uttryck krävs kunskap om de karakteristiska värdena hos dess parametrar. Bresin och Friberg [3] undersökte hur 20 musiker manipulerade sju variabler för att kommunicera fem olika emotionella uttryck. Deras resultat bekräftade att variabler som tempo, ljudnivå och artikulation konsekvent ändras i samma riktning oberoende av själva kompositionen i det undersökta materialet. Resultaten ger kvantitativa data för hur dessa variabler samverkar för att skapa en tydlig emotionell karaktär. Dessa insikter ligger till grund för de termer som används i denna studies parametriska prompts (se Tabell 1)

Tabell 1: Musikaliska parametrar för emotionell kontroll

Parameter	Definition i studien	Emotionell korrelation (exempel)
Tempo	Notdensitet mätt i antal noter per sekund.	Högre notdensitet associeras ofta med glädje, ilska och rädsla.
Ljudstyrka	Amplitud och dynamik mätt i decibel (dB).	Hög aktivitet (till exempel glädje) korrelerar med högre ljudnivå.
Artikulation	Graden av separation mellan noter (legato till staccato).	Staccato associeras med glädje/aktivitet; legato med sorg/lugn.
Register	Det frekvensområde (tonhöjd) där musiken framförs.	Högt register är typiskt för glädje; lågt för sorg.
Attack-hastighet	Hur snabbt en tons anslag når sin fulla styrka.	Snabb attack associeras med aktivitet; långsam med lugn/sorg.
Instrumentering	Val av instrument baserat på deras naturliga klangfärg (timbre/brightness).	Ljus klangfärg (t.ex. trumpet) för glädje; mörkare (t.ex. valthorn) för sorg/rädsla.

2.2 Musikens roll i audiovisuella sammanhang

I audiovisuella medier samverkar bilden ofta med en komplex ljudmiljö som kan bestå av musik, tal och ljudeffekter. Cohen [6] beskriver hur musik kan fungera som en emotionell ram som styr publikens tolkning av visuellt material. Chion [5] introducerar begreppet mervärde (*added value*), vilket innebär att ljud tillför uttryck och mening till bilden som inte finns i det visuella materialet i sig.

Forskning visar att samma visuella scen kan tolkas radikalt olika beroende på dess soundtrack. Vitouch [17] demonstrerade detta genom ett experiment där deltagarna fick se inledningsscenen från filmen *The Lost Weekend* (1945). Scenen börjar med en panoramavy över New York som sedan rör sig längs en husvägg och in genom ett öppet fönster till huvudkaraktärens lägenhet. Deltagarna fick se klippet med antingen originalmusiken (en dramatisk komposition av Miklós Rózsa) eller Samuel Barber's ‘’*Adagio for Strings”* (mer sorgsen och lågmäld). Resultatet visade att musiken styrde deltagarnas förväntningar på handlingen då de som hörde originalmusiken väntade sig en spänningsfylld fortsättning, medan de som hörde Barber tolkade scenen som betydligt mer sorgsen och vemodig.

Dessa kognitiva effekter har utforskats vidare av Boltz [2], som undersökte hur musik påverkar vårt minne av visuella händelser. I Boltz studie fick deltagarna se tvetydiga filmklipp som parades ihop med antingen positiv musik, negativ musik eller ingen musik alls. Direkt efteråt fick deltagarna beskriva karaktärernas personlighet och förutspå hur filmen skulle sluta. Deras svar följde konsekvent musikens stämning där positiv musik ledde till optimistiska tolkningar av karaktärernas motiv, medan negativ musik skapade misstänksamhet. En vecka senare genomfördes även ett minnestest som visade att deltagarna kom ihåg detaljer och föremål från scenerna mycket bättre när musiken de hört matchade scenens känsla. Boltz slutsats var att musiken fungerar som ett narrativt schema, som hjälper hjärnan att välja ut och organisera vilken information från bilden som ska lagras i minnet.

2.3 Visuell mångtydighet och emotionell projektion

Enligt Boltz [2] utövar musik sitt största inflytande i scener som besitter en hög grad av narrativ mångtydighet. Även Vitouch [17] använde en scen som var avsiktligt öppen för att demonstrera att musiken påverkar förväntningarna på vad som skulle hända härnäst.

Cohen [6] förklarar genom sin *Congruence-Association Model*, att åskådaren söker efter strukturella matchningar mellan ljud och bild. När musiken och bilden uppvisar kongruens, påverkar musikens affektiva karaktär den visuella tolkningen. Detta gör att en scen kan ändras från exempelvis sorgsen till hoppfull enbart genom ljudspårets inverkan.

För denna studie är detta perspektiv avgörande, då användandet av ett narrativt mångtydigt stimuli skapar de förutsättningar som krävs för att isolera musikens emotionella effekt. Genom att välja ett visuellt material som inte har en stark inneboende känsla (se 3.2 & 3.2.1), kan studien undersöka AI-musikens förmåga att på egen hand styra den emotionella tolkningen.

2.4 Generativ musik och styrbarhet i AI-system

Traditionellt musikskapande för audiovisuella medier har historiskt sett varit en process som ofta kräver djupgående kunskaper i musikteori, hantering av avancerad mjukvara samt tillgång till professionella kompositörer. Modeller för generativ musik som Suno och Udio har sänkt tröskeln för musikkompositioner i och med möjligheten att snabbt generera musik med hjälp av språkliga instruktioner. Som tidigare nämnts i introduktionen (se avsnitt 1) finns det både kritik kring den etiska aspekten av AI-genererad musik samt problematik kring bristande insyn i interaktionen med det generativa systemet. Att undersöka systemen blir därmed också en fråga kring att kritiskt granska om de faktiskt är tillförlitliga bärare av mänsklig intention.

Forskning kring hur väl dessa system förmedlar avsedda känslor pekar på betydande utmaningar. En nyligen gjord studie av Hu [11] undersökte hur den avsedda känslan i textprompter reflekteras i den genererade musiken genom att analysera över 6000 par av prompter och ljudfiler insamlade från Suno och Udio. Resultaten visade att det fanns en signifikant men svag korrelation mellan känslomässiga värden i prompter och den resulterande musiken. Dessutom konstaterade Hu att AI-modeller har lättare att fånga och återge arousal än valens, samt att bristande överensstämmelse ofta beror på neutralt språk eller motstridiga instruktioner i användarnas prompter. En viktig aspekt var också inkluderingen av lyrik, vilket visade sig ha en betydande positiv inverkan på AI-modellernas förmåga att förutsäga musikens valens. Hu påpekar dock att närvaron av lyrik kan göra det svårt att avgöra hur mycket av den emotionella precisionen som faktiskt beror på de rent
musikaliska egenskaperna kontra de semantiska ledtrådarna i texten.

Ytterligare forskning har gjorts av Fišer et al. [7] som undersökt den känslomässiga effekten av AI-genererad kontra mänskligt komponerad musik i audiovisuella medier. I studien användes två olika typer av strategier för AI-musiken, simpla instruktioner med mindre detaljer och en med mer tekniska och detaljerade instruktioner. Studien gjorde dels biometriska mätningar och konstaterade att båda

typerna av AI-genererad musik kräver högre kognitiv ansträngning för lyssnaren att bearbeta jämfört med mänskligt skapad musik. Deras resultat visade att AI-musik framkallade större pupillvidgning, vilket tyder på ökad mental belastning, samt upplevdes som mer arousal-höjande men mindre bekant än mänsklig musik. AI-musiken med mer detaljerade prompts resulterade även i högre blinkhastighet och hudimpedans som är markörer för högre uppmärksamhet och kognitiv belastning. Studien visade ingen signifikant skillnad i resultat av skattad valens mellan de tre olika kompositionerna, men betonar att en viktig aspekt att ta hänsyn till är den mänskliga synens dominans över hörseln, som kan påverka resultaten i audiovisuella sammanhang.

2.5 Definitioner

I detta arbete definieras följande begrepp:

- **Semantisk emotionell prompting:** Instruktioner som beskriver ett önskat känsloläge (t.ex. "lugn", "spänning") utan att ange de musikaliska medel som krävs för att nå dit.
- **Strukturell parametrisk prompting**: Instruktioner som beskriver specifika musikaliska och tekniska egenskaper (se Tabell 1)

2.6 Forskningsgap

Då Hu [11] pekar på att det finns en signifikant men svag korrelation mellan känslomässiga värden i prompter och den resulterande musiken undersöker denna studie om mer musikteoretiska specificerade promptar kan öka denna korrelation.

Vidare kompletteras forskningen av Fišer et al. [7] genom en förändrad audiovisuell metodik. Medan deras studie använde videoklipp med starka narrativa betydelser (såsom djurdokumentärer eller spänningsscener), använder detta arbete endast ett material i en mer mångtydig visuell miljö. Syftet är att isolera musikens inverkan och förhindra att bildens innehåll i sig själv påverkar den emotionella intentionen.

3. Metod

Denna studie genomfördes som en experimentell undersökning med en inomgruppsdesign, vilket innebar att samtliga deltagare exponerades för alla sex betingelser. För att isolera musikens förmåga att förmedla emotionell intention användes ett specifikt visuellt klipp som fungerade som en fast referenspunkt genom hela experimentet. Syftet med det visuella materialet var att använda ett material som var öppet för tolkning för att tillåta att musiken aktivt kunde styra den emotionella tolkningen (se 3.2 & 3.2.1 för detaljerad beskrivning av det visuella stimulit).

3.1 Variabler

För att svara på forskningsfrågan definierades följande variabler:

Oberoende variabel (IV): Promptingstrategi. Denna manipulerades i två nivåer: *semantisk emotionell prompting* (abstrakta instruktioner) och *strukturell parametrisk prompting* (teoretiskt specifika instruktioner).

Beroende variabler (DV): För att få en helhetsbild av hur musiken uppfattas och hur väl den matchar den ursprungliga intentionen, mättes tre beroende variabler. Dessa valdes för att ge både en matematisk precision, en kategorisk tydlighet och en djupare kvalitativ förståelse:

1. **Emotionell precision:** Detta mättes genom att deltagarna skattade musiken utifrån dimensionerna valens och arousal (se 2.1). Precisionen beräknades som det euklidiska

avståndet mellan deltagarnas skattningar och de fördefinierade målkoordinaterna (se 3.2.1) för varje känsla. Denna variabel är central för att svara på forskningsfrågan då den ger ett objektivt mått på hur nära AI:ns output hamnar det avsedda målet i det emotionella rummet (se avsnitt 3.6 för matematisk formel).

2. **Kategorisk matchning:** Detta mättes genom att deltagarna fick identifiera den mest framträdande känslan ur en lista bestående av de tre målkategorierna (lugn, spänning, sorg) samt alternativet ”Övrigt” med möjlighet till fritext. Denna variabel valdes för att mäta hur entydigt musiken tolkas av lyssnaren. Genom att inkludera alternativet ”Övrigt” skapades även utrymme att fånga upp närliggande känslolägen som faller inom samma emotionella fält som målkategorin, även om deltagaren inte valde exakt den föreslagna etiketten.
3. **Subjektiv interpretation (Kvalitativ):** Genom frivilliga textfrågor fick deltagarna beskriva sina upplevelser med egna ord. Denna variabel valdes för att fånga upp de nyanser och narrativa associationer som de numeriska skalorna missar.

3.2 Stimuli: Visuellt material

För att undersöka den emotionella precisionen av musiken valdes en 30 sekunder lång videosekvens från videobanken Pexels[3]. Videon[4] består av en panoramavy över ett bergigt landskap där ett tåg färdas över en bro och slutligen försvinner in i en tunnel (se Figur 2). Klippets originalljud har även tagits bort vid utförandet av experimentet för att endast den genererade musiken ska höras. Klippet valdes på grund av dess narrativa mångtydighet (se 2.3). Även om en panoramavy över ett landskap med ett tåg bär på vissa inneboende associationer, saknar scenen en fastställd emotionell riktning eller mänsklig protagonist. Detta gör klippet lämpligt som visuellt stimuli eftersom musiken får fungera som den primära meningsbäraren. Kameratekniskt karaktäriseras klippet av en långsam, svepande kamerarörelse som kan upplevas ha en slow-motion-liknande karaktär.

Figur 2: *Stillbild från det visuella stimulit föreställande ett tåg i ett bergigt landskap.*

[3] https://www.pexels.com/

[4] https://www.pexels.com/sv-se/video/naturskon-schweizisk-tagoverfart-over-majestatisk-viadukt-32817463/

3.2.1 Validering genom pilotstudie

För att ytterligare stärka klippets lämplighet som visuellt stimuli utfördes en pilotstudie med ($n = 7$) deltagare. Deltagarna bestod av 3 kvinnor och 4 män med en medelålder på ($M = 35,9$) och en standardavvikelse på ($SD = 19,1$). Deltagarna fick titta på videosekvensen utan ljud och därefter skatta sin upplevelse utifrån studiens skalor för valens och arousal (1–7). Detta gjordes genom två frågor: dels gällande **emotionell valens** (”Hur upplevde du stämningen i klippet?”), där 1 motsvarade mycket negativ/obehaglig och 7 mycket positiv/behaglig, och dels gällande **emotionell arousal** (”Hur skulle du beskriva energinivån i klippet?”), där 1 motsvarade mycket låg/lugn och 7 mycket hög/spänd. Resultatet från pilotstudien visade ett medelvärde på **4,29** för valens ($SD = 0,76$) och **3,43** för arousal ($SD = 0,53$). Då båda medelvärdena ligger nära skalans neutrala mittpunkt (4,0) och uppvisar en låg spridning styrker det klippets lämplighet som ett emotionellt öppet stimuli. Detta gjorde att deltagarnas skattningar i huvudstudien med större sannolikhet kan tillskrivas musikens inverkan snarare än emotionella egenskaper i det visuella materialet i sig.

3.3 Generering av musik

Musiken genererades med hjälp av Suno v4.5, som är den senaste modellen tillgänglig för gratisanvändare vid tidpunkten för denna studie. Vid generering i Suno v4.5 skapas alltid två alternativa kompositioner baserat på samma prompt. I denna studie gjordes ett manuellt val kring vilken version som bedömdes ha störst potential att förmedla den avsedda känslan. Därefter valdes även det mest representativa segmentet om 30 sekunder ut från den valda versionen. Samtliga prompts formulerades på engelska för att minimera risken för tolkningsfel. För de semantiska emotionella prompterna användes ett så enkelt och tydligt språk som möjligt, där endast den avsedda känslan specificerades. De strukturella parametriska promptsen utgick från Tabell 1 (se 2.1) men med tilläggen orkestral stil samt specificerad tonart. Valet av att endast använda orkestral musik motiverades av att att båda promptingsstrategierna skulle falla inom samma musikgenre, men även den orkestrala musikens etablerade roll som det primära tonspråket inom västerländsk film [13].

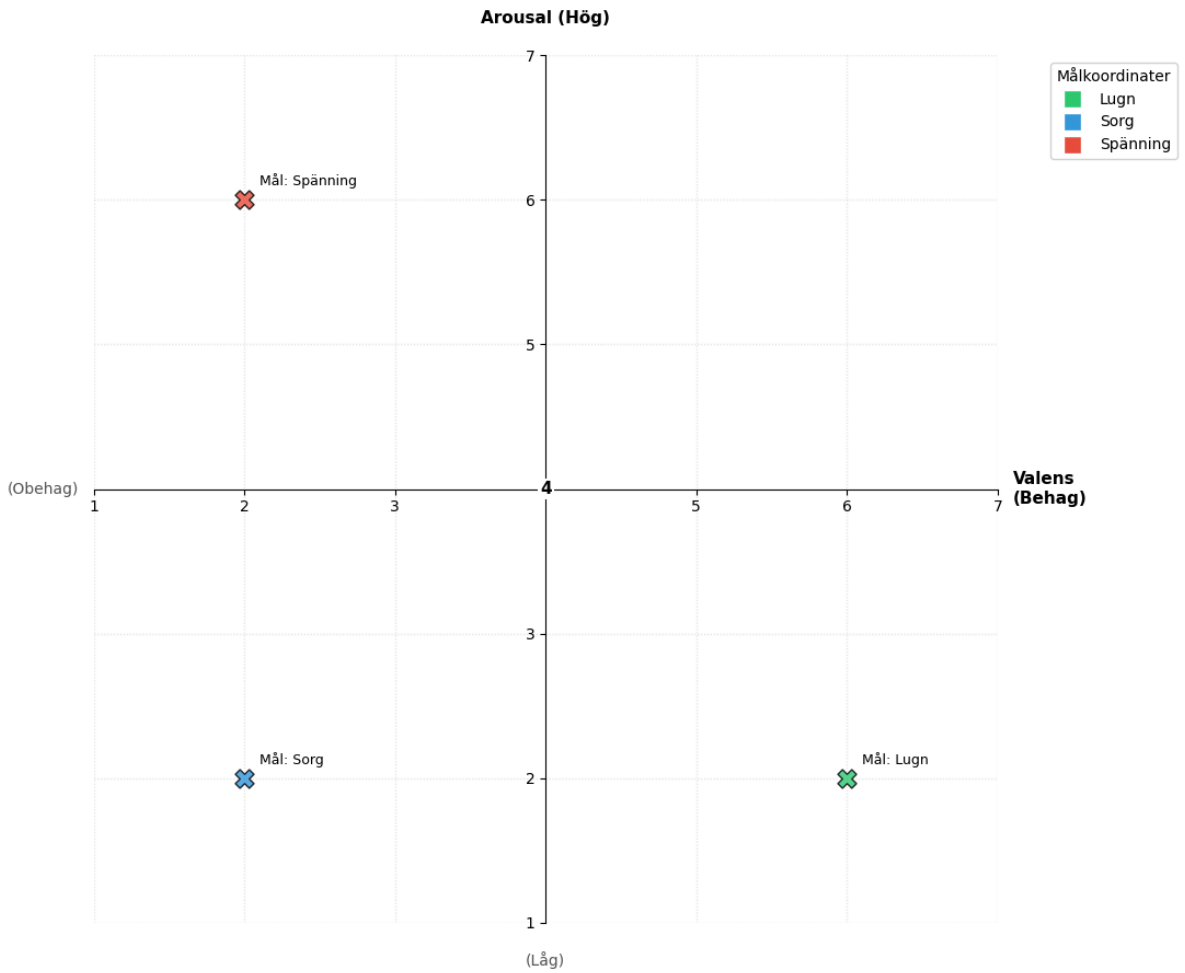
Tabell 2: Operationalisering av betingelser och prompts i Suno v4.5

*Notera: Samtliga länkar är klickbara. QR-koder och fullständiga länkar till materialet finns tillgängliga i **Bilaga 1**.*

Avsedd Emotion	Målkoordinater (Valens, Arousal)	Semantisk emotionell prompting (Abstrakt)	Strukturell parametrisk prompting (Konkret)
Lugn (Calm)	V: 6, A: 2	"Make an orchestral piece that conveys calmness."	"Orchestral, C Major, low note density, low amplitude, legato articulation, mid register, slow attack speed, warm instrumentation like cellos and woodwinds."
Spänning (Tension)	V: 2, A: 6	"Make an orchestral piece that conveys tension."	"Orchestral, D Minor with dissonant harmonies, high note density, high amplitude, staccato articulation, wide register with piercing high brass, fast attack speed, bright instrumentation."
Sorg (Sorrow)	V: 2, A: 2	"Make an orchestral piece that conveys sorrow."	"Orchestral, G Minor, very low note density, low amplitude, legato articulation, low register, very slow attack speed, dark and mellow.'' instrumentation like double basses and French horns."

3.3.1 Motivering av målkoordinaterna

Målkoordinaterna för de tre emotionella kategorierna utgick från Russells [16] modell (se Figur 1). Specifika emotionella etiketter (såsom spänning eller sorg) kan variera något mellan olika tolkningar i modellen. Ibland kan exempelvis spänning placeras med en något högre valens och även sorg kan ha varierande hög arousal. Men konsekvent för de olika känsloetiketterna är att de alltid placeras i samma kvadrant oavsett små skillnader av tolkningar för modellen. För denna studie fattades ett metodologiskt beslut att standardisera dessa positioner för att säkerställa experimentell symmetri.



Figur 3: Visualisering av målkoordinaterna i Valens-Arousal-rymden

Målpunkterna placerades (se Figur 3) enligt följande koordinater på den använda Likert-skalan (1–7), där 4 representerar neutralt mittläge:

- **Lugn:** $(V = 6, A = 2)$— Hög valens, låg arousal.
- **Sorg:** $(V = 2, A = 2)$ — Låg valens, låg arousal.
- **Spänning:** $(V = 2, A = 6)$ — Låg valens, hög arousal.

Valet av dessa specifika punkter motiverades av tre faktorer:

1. **Experimentell balans (Euklidisk symmetri):** Samtliga punkter är placerade med exakt samma avstånd från skalans mittpunkt $(V = 4, A = 4)$. Varje punkt ligger 2 enheter horisontellt och 2 enheter vertikalt från centrum. Detta innebär att det euklidiska avståndet till målet blir matematiskt jämförbart mellan alla klipp, vilket förhindrar att en kategori får en fördel eller nackdel i precisionsmätningar baserat på dess placering i diagrammet.
2. **Kategorisk distinktion:** Genom att placera punkterna i motsatta hörn av de valda kvadranterna maximeras det emotionella avståndet. Detta minskar risken för överlappning och skapar tre distinkta målkänslor.
3. **Prototypiska emotioner:** Placering av målkoordinaterna representerar den tydligaste tolkningen av emotionen. Även om känsloetiketterna kan variera något är dessa känslor generellt förknippade med extremvärdena inom valens respektive arousal.

3.4 Datainsamling och procedur

Data samlades in genom en digital enkät gjord i Google Forms. För att minska ordningseffekterna skapades tre enkäter (A, B och C) med olika ordning på klippen. Deltagarna tilldelades slumpmässigt en av dessa versioner. Den specifika ordningsföljden för respektive enkätversion presenteras i Tabell 3.

Tabell 3. Stimuli-ordning för de tre enkätversionerna.

Position	Enkät A	Enkät B	Enkät C
1	Semantisk Lugn	Strukturell Sorg	Strukturell Spänning
2	Strukturell Spänning	Strukturell Lugn	Semantisk Sorg
3	Strukturell Sorg	Semantisk Spänning	Semantisk Lugn
4	Strukturell Lugn	Semantisk Sorg	Semantisk Spänning
5	Semantisk Spänning	Semantisk Lugn	Strukturell Sorg
6	Semantisk Sorg	Strukturell Spänning	Strukturell Lugn

Vid utformningen av de tre enkätversionerna tillämpades en systematisk motbalansering av stimuli-materialet. För att kontrollera för ordningseffekter utformades ordningsföljden utifrån två kriterier:

1. **Varierad startpunkt:** Varje enkätversion inleddes med ett klipp från en unik känslokategori (Lugn, Sorg respektive Spänning).
2. **Eliminering av kategorisk upprepning:** Ordningsföljden arrangerades så att inga stimuli från samma känslokategori presenterades efter varandra.

Denna struktur syftade till att minimera risken för att deltagarnas bedömningar skulle påverkas av en specifik inledande känsla. Att samma känsla inte presenterades direkt efter varandra syftade till att förhindra att deltagarna inte skulle göra jämförelser mellan de två olika klippen istället för att skatta dem individuellt.

Efter varje klipp besvarade deltagarna följande fyra frågor:

1. **Emotionell valens:** "Hur upplevde du stämningen i klippet?" (1: Mycket negativ/obehaglig — 7: Mycket positiv/behaglig).
2. **Emotionell arousal:** "Hur skulle du beskriva intensiteten/energinivån i klippet?" (1: Mycket låg/lugn — 7: Mycket hög/spänd).
3. **Kategorisk matchning:** "Vilket av följande ord beskriver bäst klippets stämning?" Deltagarna fick välja mellan alternativen "Lugn", "Spänning", "Sorg", eller "Övrigt", (där de själva fick skriva ett beskrivande ord).

4. **Frivillig reflektion** - frivillig textinput där deltagaren har möjlighet att beskriva hur de upplevde varje klipp.

I slutet av enkäten fanns även möjligheten för en fri reflektion kring hela experimentet. Frågan löd: "Är det något mer du vill tillägga/reflektera över efter att ha sett alla 6 stycken klipp?".

3.5 Deltagare och urval

Urvalet för studien bestod av 31 deltagare. Rekryteringen skedde genom ett snöbollsurval, där studien spreds via personliga nätverk där deltagarna sedan uppmuntrades att i sin tur dela enkäten vidare. Detta resulterade i ett urval med en medelålder på ($M = 35,77$) och en standardavvikelse på ($SD = 13,93$). Av samtliga deltagare uppgav 21 personer att de hade ingen erfarenhet av musikproduktion, 7 personer uppgav att de var på en hobbynivå och 3 personer uppgav att de höll på professionellt med musikskapande. Fördelningen för urvalet var 14 män och 17 kvinnor.

En viktig metodologisk aspekt är att deltagarna under testmomentet inte var medvetna om att musiken var genererad av AI. Detta val gjordes för att eliminera risken för fördomar eller en negativ attityd mot AI-genererad musik, då detta skulle kunna påverka den emotionella upplevelsen och skattningarna.

Samtliga deltagare informerades om att deras medverkan var helt frivillig och att de när som helst kunde avbryta sitt deltagande. Inga känsliga personuppgifter samlades in utöver grundläggande demografi som kön och musikalisk erfarenhet.

3.6 Analysmetod

Kvantitativ analys: Som ett första steg i analysen beräknades medelvärden (M) och standardavvikelser (SD) för deltagarnas skattningar av valens och arousal för varje klipp. För att möjliggöra en visuell utvärdering av resultaten placerades dessa medelvärden i ett spridningsdiagram representerande valens-arousal-rymden. I detta diagram inkluderades även studiens teoretiska målkoordinater för respektive känsla (lugn, sorg och spänning), vilket skapade en grund för att identifiera generella mönster och avvikelser i hur väl de olika klippen kommunicerade den avsedda känslan.

Mätningen av emotionell precision beräknades som det euklidiska avståndet d mellan deltagarens skattning (v_1, a_1) och målkoordinaten (v_2, a_2) enligt formeln:

$$d = \sqrt{(v_1 - v_2)^2 + (a_1 - a_2)^2}$$

Ett lägre värde på d tolkades som en högre grad av precision.

För att jämföra de två promptingstrategierna genomfördes parvisa t-test i Google Sheets. Valet av ett parametriskt test motiverades av studiens inomgruppsdesign, där samma deltagare utvärderade båda strategierna, samt att det euklidiska avståndet utgjorde kontinuerlig data på en intervallskala. Då urvalsstorleken uppgick till 31 deltagare ($n = 31$) kunde datan antas vara normalfördelad enligt centrala gränsvärdessatsen. En signifikansnivå på $p < 0,05$ användes för att fastställa statistiskt säkerställda skillnader.

Kategorisering: Träffsäkerheten analyserades som ett komplement till den emotionella precision genom "direkt matchning" (exakt målord) och "bred matchning". Den breda matchningen definierades som svar som överensstämde med målets emotionella valens (positiv (5-7) eller negativ (1-3)) riktning, oavsett intensitetsnivå. Resultaten sammanställdes procentuellt för att ge en överskådlig bild av hur ofta den övergripande känslan prickades rätt.

Kvalitativ analys: Deltagarnas frivilliga reflektioner analyserades genom en förenklad tematisk analys. Processen innebar att svaren lästes igenom upprepade gånger för att identifiera återkommande teman och mönster i hur deltagarna tolkade musiken och bildens inverkan på upplevelsen. Svaren grupperades utifrån de tre känslokategorierna, samt en sammanställning av kvalitativa svar kring experimentet i sin helhet.

4. Resultat

I denna del presenteras studiens resultat. Resultatet börjar med en analys kring deltagarnas skattning av valens och arousal och visualiserar dessa i valens-arousal-rymden. Vidare analyseras den emotionella precisionen i form av det euklidiska avståndet till målkoordinaterna för respektive stimuli. Resultatet tar sedan upp hur deltagarna kategoriserat de olika klippen och presenterar även de frivilliga textsvaren i kategorin ”Övrigt” och både en direkt och bred matchning presenteras. Slutligen presenteras de frivilliga reflektionerna kring varje klipp samt övergripande reflektioner kring hela experimentet.

4.1 Skattning av Valens och Arousal

I tabell 4 redovisas medelvärdet (M) och standardavvikelsen (SD) för deltagarnas skattningar av valens och arousal för de sex olika klippen.

Tabell 4: Övergripande emotionell skattning

Stimuli	Valens (M)	Valens (SD)	Arousal (M)	Arousal (SD)
Semantisk Lugn	5,94	0,96	3,90	1,62
Strukturell Lugn	5,45	1,31	3,39	1,20
Semantisk Sorg	4,58	1,38	3,03	0,98
Strukturell Sorg	5,06	1,43	3,55	1,26
Semantisk Spänning	3,32	1,05	5,29	0,78
Strukturell Spänning	2,97	1,22	6,35	0,61

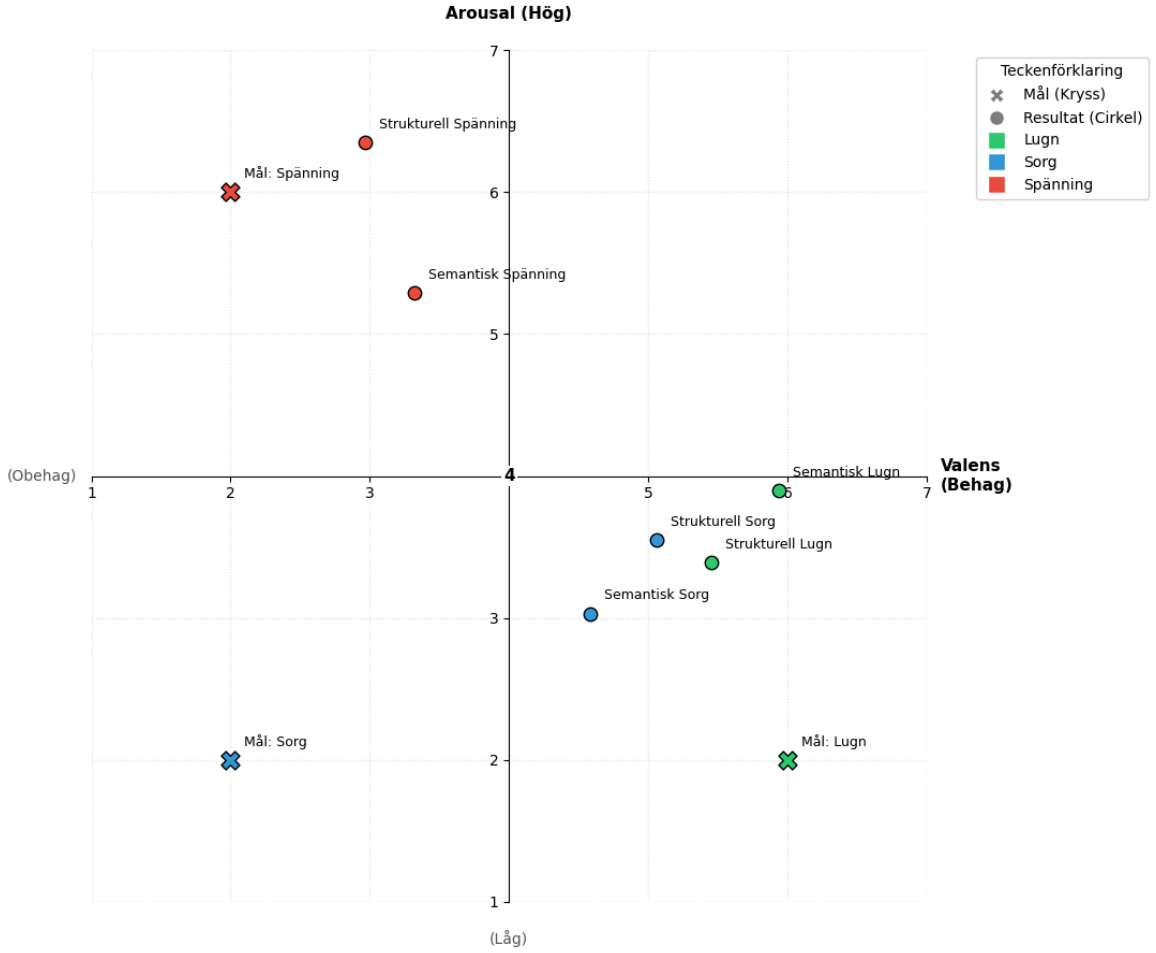
Tabellen visar att klippen för kategorin *Lugn* erhöll de högsta valensvärdena, där den semantiska strategin skattades högst $(M = 5,94)$. För kategorin *Spänning* genererade den strukturella strategin den lägsta valensskattningen $(M = 2,97)$ och den högsta arousalnivån $(M = 6,35)$. Noterbart är

att skattningarna för *Sorg* visar att båda strategierna resulterade i valensmedelvärden över skalans mittpunkt (4, 58 respektive 5, 06), trots ett emotionellt mål på (2, 0).

Vidare kan noteras att den semantiska strategin för *Sorg* resulterade i en lägre arousalskattning ($M = 3,03$) jämfört med den strukturella ($M = 3,55$). Standardavvikelserna var generellt lägst för kategorin spänning, vilket kan tyda på en högre samstämmighet bland deltagarna för dessa stimuli jämfört med *Lugn* och *Sorg*.

4.1.1 Placering i Valens-Arousal-rymden

För att visualisera hur de övergripande skattningarna förhåller sig till studiens målkoordinater har medelvärdena prickats in i ett spridningsdiagram (se Figur 4) över Russells [16] modell (se Figur 2). Diagrammet är uppdelat i fyra emotionella kvadranter av axlar som korsar vid skalans mittpunkt (4, 4). Målkoordinaterna för de tre känslorna är markerade med ett kryss, och medelvärdena av deltagarnas skattningar med en cirkel.



Figur 4: Diagram över valens-arousal-rymden med utplaceringar av målkoordinater och resultatet för de olika stimulin.

Som framgår av Figur 4 varierar precisionen mellan de olika känslokategorierna och strategierna. För kategorin *Lugn* (gröna markörer), med målkoordinat (6, 2), återfinns båda resultaten i rätt kvadrant (nedre högra; hög valens, låg arousal). Skattningarna för valens är mycket nära målet, där *Semantisk*

Lugn har en nästan exakt matchning. Arousal-skattningarna är dock förhöjda jämfört med målet, särskilt för *Semantisk Lugn*. För kategorin *Spänning* (röda markörer), med målkoordinat $(2, 6)$, är båda resultaten väl placerade i den avsedda kvadranten (övre vänstra; låg valens, hög arousal), med relativt god precision även om valensskattningarna är något högre än målet. Den största och mest konsekventa avvikelsen ses för kategorin *Sorg* (blå markörer). Målkoordinaten $(2, 2)$ ligger i den nedre vänstra kvadranten (låg valens, låg arousal). I stark motsats härtill återfinns båda resultaten, *Strukturell Sorg* och *Semantisk Sorg*, i den nedre högra kvadranten, som representerar hög valens och låg arousal. *Semantisk Sorg* uppvisar en bättre precision på båda axlarna jämfört med *Strukturell Sorg*.

4.2 Statistisk analys av emotionell precision

För att testa om det fanns en statistisk signifikant skillnad i precision mellan den semantiska och den strukturella promptingstrategin genomfördes parvisa t-test. Precisionen definierades som det euklidiska avståndet mellan deltagarnas individuella skattningar och klippets målkoordinat i valens-arousal-rymden. Ett lägre värde betyder alltså ett mindre avstånd och därmed bättre precision. Resultaten av dessa tester, tillsammans med medelvärden M och standardavvikelser SD för avstånden, presenteras i Tabell 5.

Tabell 5. Emotionell precision (euklidiskt avstånd) och resultat av t-test.

Känslomål	Semantisk Precision (M)	Semantisk Precision (SD)	Strukturell Precision (M)	Strukturell Precision (SD)	p-värde
Lugn	2,17	1,55	2,05	1,06	0,660
Sorg	2,95	1,37	3,74	1,15	**0,041***
Spänning	1,70	1,02	1,40	0,96	0,126

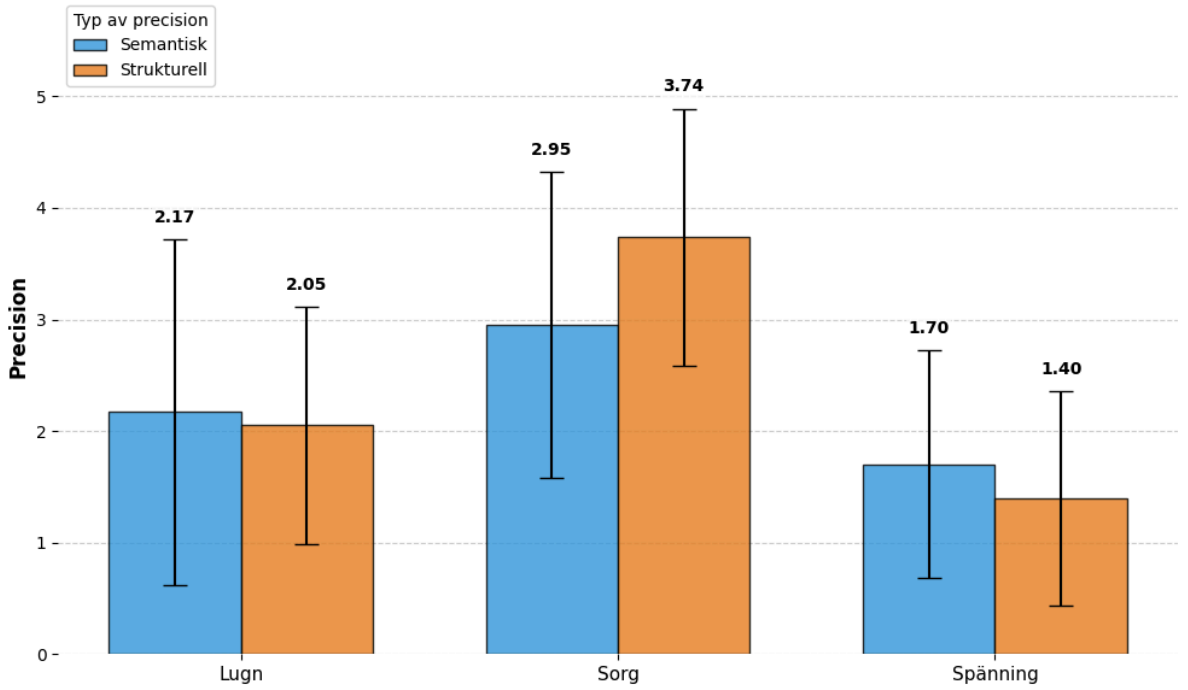
**Indikerar statistisk signifikans* ($p < 0,05$)

Som framgår av Tabell 5 uppvisade kategorin *Sorg* en statistiskt signifikant skillnad mellan de två strategierna ($t(30) = -2.13$; $p = 0,041$). Den semantiska strategin resulterade i ett signifikant lägre euklidiskt avstånd ($M = 2,95$) jämfört med den strukturella strategin ($M = 3,74$), vilket indikerar att den semantiska prompten lyckades förmedla känslan av sorg med högre precision.

För kategorierna *Lugn* ($p = 0,660$) och *Spänning* ($p = 0,126$) återfanns inga statistiskt signifikanta skillnader mellan strategierna, även om spänning visar på en tendens till en skillnad. Noterbart är dock att den högsta generella precisionen (lägst medelvärde för avstånd) återfanns för kategorin *Spänning* vid användning av den strukturella strategin ($M = 1,40$).

Standardavvikelserna för precisionen var generellt högre för de klipp som syftade till att förmedla sorg och lugn jämfört med spänning, vilket tyder på en större individuell variation i hur deltagarna uppfattade dessa stimuli i förhållande till de uppsatta målen.

För att tydliggöra skillnaderna i precision mellan de två promptingstrategierna visualiseras medelvärdena för det euklidiska avståndet i Figur 5. Diagrammet inkluderar även felstaplar som representerar standardavvikelsen, vilket ger en bild av spridningen i deltagarnas skattningar för respektive klipp.



Figur 5. Emotionell precision per kategori med standardavvikelse.

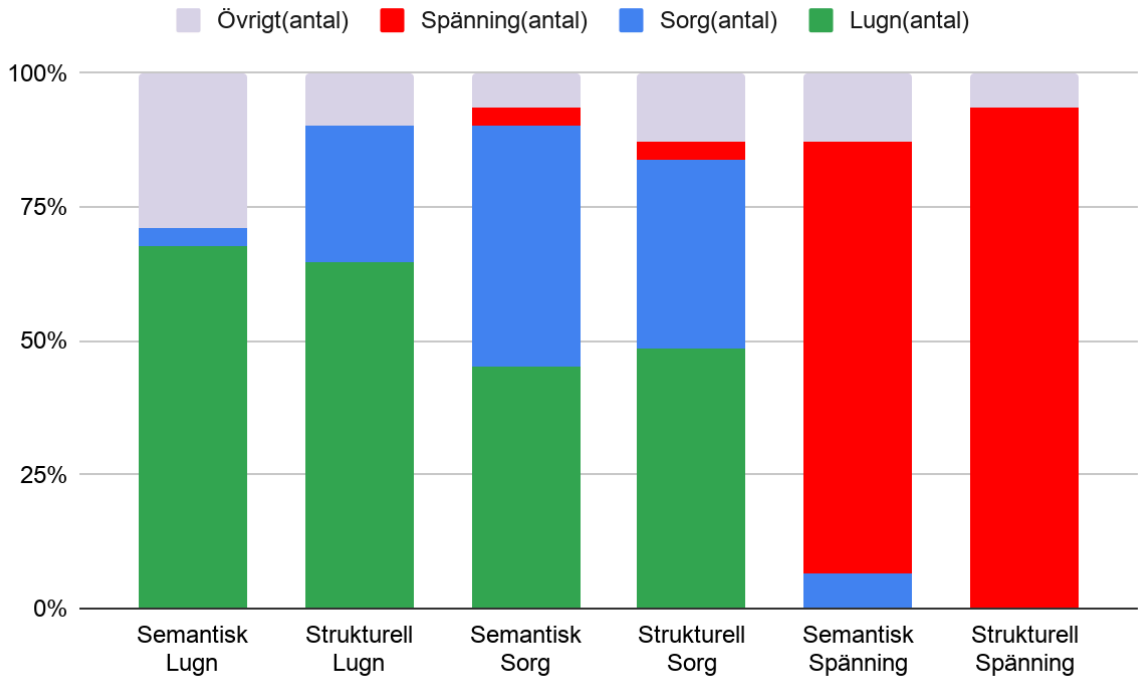
Som framgår av Figur 5 är det euklidiska avståndet konsekvent lägre för kategorin *Spänning*, vilket indikerar att båda strategierna lyckades förmedla denna känsla med högst precision. Den mest framträdande visuella skillnaden återfinns i kategorin *Sorg*, där den semantiska strategins stapel ($M = 2,95$) är tydligt lägre än den strukturella ($M = 3,74$). Detta illustrerar grafiskt det signifikanta p-värde som redovisades tidigare i Tabell 5.

Felstaplarna visar på en betydande spridning i resultaten, särskilt för kategorierna *Lugn* och *Sorg*. Den stora spridningen tyder på att deltagarnas uppfattning av dessa klipp varierade mer i förhållande till målet än vad som var fallet för spänningsklippen, där felstaplarna är något kortare.

För *Lugn* och *Spänning* är staplarna för den emotionella precisionen relativt likvärdiga i höjd, vilket illustrerar avsaknaden av statistisk signifikans för dessa kategorier.

4.3 Kategorisering

För att undersöka deltagarnas kategoriska uppfattning av de olika stimuli-materialen fick deltagarna välja det ord som bäst beskrev klippets emotionella karaktär. Resultatet presenteras i Figur 6, där varje stapel visar den procentuella fördelningen av svar för respektive klipp.



Figur 6. Fördelning av deltagarnas valda emotionella kategorier (antal i procent) per videoklipp

Som framgår av Figur 6 varierade träffsäkerheten (andelen deltagare som valde målkategorin) stort mellan de olika emotionella målen. Kategorin *Spänning* (röd) uppvisade den högsta kategoriska överensstämmelsen, där nästan samtliga deltagare valde rätt kategori för det strukturella klippet. För den semantiska strategin för *Spänning* valde 80,6% rätt ord, medan en liten andel valde "Sorg" eller "Övrigt". Medans för den strukturella strategin valdes endast "Spänning" (93,5 %) eller "Övrigt" (6,5 %).

För kategorin *Lugn* (grön) var resultaten likvärdiga mellan strategierna, med en träffsäkerhet på drygt 60 %. Noterbart är att en betydande del av deltagarna valde kategorin *Sorg* för *Strukturell Lugn*, medan en större andel valde "Övrigt" för *Semantisk Lugn*.

Den mest splittrade kategoriseringen återfinns i kategorin *Sorg* (blå). För båda strategierna valde mindre än hälften av deltagarna målkategorin *Sorg*. Istället uppfattades dessa klipp i hög grad som *Lugn* (grön). Detta korrelerar med de tidigare redovisade resultaten i Figur 4 (se 4.1.1), där klippen för *Sorg* hamnade i den positiva valens-kvadranten. Den semantiska strategin för *Sorg* hade en något lägre andel felaktiga skattningar som "Lugn" jämfört med den strukturella, men uppvisade istället en större andel svar i kategorin "Övrigt".

4.3.1 Analys av svarsalternativet ”Övrigt”

För att få en djupare förståelse för de fall där deltagarna inte ansåg att målorden var representativa, analyserades fritextsvaren för alternativet ”Övrigt”. Dessa presenteras i Tabell 6.

Tabell 6. Sammanställning av fritextsvar för alternativet ”Övrigt”.

Klipp	Egna beskrivningar (Fritext), (antal)
Semantisk Lugn	Romantisk (3), Hoppfullt, Förhöjd positiv känsla, Episk förväntan, triumf och frihet, Glädje, Äventyrlig
Semantisk Sorg	Romantisk, Förväntansfull
Semantisk Spänning	Brådska, Stress, Transportsträcka, Vet ej
Strukturell Lugn	Hoppfullt (2), Positivt
Strukturell Sorg	Episkt, Stämningsfullt, Crescendo, Kärleksfullt
Strukturell Spänning	Hektiskt, Dystert

Analysen av fritextsvaren i Tabell 6 uppvisar olika teman beroende på klipp och strategi. För *Semantisk Lugn* används termer med romantiska inslag och storslagenhet, som exempelvis ”Romantik”, ”Triumf”, ”Äventyrlig” och "Episk förväntan”. För båda strategierna inom kategorin *Sorg* användes ord med positiv laddning såsom "Romantiskt", "Kärleksfullt" och "Episkt". Inom kategorin *Spänning* rörde beskrivningarna främst tidsaspekter och stress, exempelvis ”Brådska” och ”Hektiskt”.

4.3.2 Analys av träffsäkerhet

Som ett slutgiltigt mått på hur väl de emotionella målen kommuniceras beräknades träffsäkerheten i form av direkt och bred matchning (se Tabell 7). Direkt matchning definieras som andelen deltagare som valde exakt målord, medan bred matchning inkluderar de fritextsvar som föll inom samma emotionella valens (positiv (5-7), negativ (1-3)) som målet. Som framgår av Tabell 7 uppvisade kategorin *Spänning* den högsta träffsäkerheten totalt sett. Särskilt utmärkte sig den strukturella strategin för spänning med en direkt matchning på 93,5 % och en bred matchning på 96,8 %.

För kategorin *Lugn* syns en tydlig skillnad mellan direkt och bred matchning, framför allt för den semantiska strategin. Trots en direkt matchning på 67,8 % ökade den breda matchningen till 96,8 %. Detta indikerar att nästan samtliga deltagare uppfattade den avsedda positiva valensen, även om de valde ord som beskriver en högre arousal (energinivå) än vad etiketten lugn förmedlar, såsom "Triumf" eller "Äventyrlig".

Kategorin *Sorg* uppvisade de lägsta värdena i studien (45,2 % respektive 48,3 %). En kritisk observation är att det för *Sorg* råder total avsaknad av skillnad mellan direkt och bred matchning. Detta innebär att ingen av de deltagare som valde svarsalternativet "Övrigt" angav ord med negativ valens.

Tabell 7. Procentuell fördelning av direkt och bred matchning per stimuli.

Klipp	Direkt matchning(%)	Bred matchning (%)
Semantisk Lugn	67,8 %	96,8%
Semantisk Sorg	45,2 %	45,2 %
Semantisk Spänning	80,1%	87,1%
Strukturell Lugn	64,5%	74,2%
Strukturell Sorg	48,3%	48,3%
Strukturell Spänning	93,5%	96,8%

4.4 Kvalitativa observationer och deltagarreflektioner

Deltagarnas frivilliga reflektioner användes för att få en djupare förståelse för den emotionella upplevelsen. Svaren har kategoriserats utifrån de tre målkänslorna samt generella kommentarer om experimentet.

4.4.2 Emotionell respons per kategori

Kategorin Spänning: Reflektionerna kring spänningsklippen präglades av en tydlig narrativ uppbyggnad där musiken gav bilden ett hotfullt eller actionorienterat sammanhang. Flera deltagare beskrev känslor av fara:

- *”Det känns som nu kommer tåget spåra ut eller nåt för boven har riggat bomber på spåren.”*
- *”Det här var intensivt, jag blev nästan rädd att tåget skulle ramla av bron.”*
- *”Hetsig låt men långsamt filmat. Det krockar = obehagligt.”*

Kategorin Lugn: För de klipp som syftade till att förmedla lugn fokuserade kommentarerna på harmoni och positiva framtidsutsikter. Här återfanns även kopplingar till filmiska avslutningsscener:

- *”Ett lugn. Fridfullt. Som en resa framåt mot bättre tider och positivitet.”*
- *”Känns som en 'end scene' i typ Harry Potter.”*
- *”Det här gav mig en fin känsla i kroppen och jag blev lugn.”*

Kategorin Sorg: Sorg-kategorin uppvisade de mest varierade reflektionerna. Medan vissa tolkade musiken som tragisk, upplevde andra den som början på något hoppfullt eller episkt:

- *”Nu dör någon huvudkaraktär kändes det som.”*

- *”Lite illavarslande, som om något sorgligt kommer att hända!”*
- *”Fick en känsla av att det var starten på ett äventyr, på något episkt.”*
- *”Kändes som en avslutning på en film, snarare än ett intro eller en scen mitt i.*”

4.4.3 Övergripande reflektioner kring musikens påverkan

I de avslutande kommentarerna rådde en samstämmighet kring musikens dominanta roll för tolkningen av det visuella materialet. Trots att bildmaterialet var identiskt i samtliga klipp, upplevde deltagarna att musiken hade en stor påverkan på scenens innebörd.

En deltagare summerade upplevelsen: *”Samma tåg, men många olika stämningar som förmedlades tack vare de olika låtarna!”*. Andra poängterade att musiken inte bara förändrade den nuvarande känslan utan även skapade förväntningar på vad som skulle hända härnäst: *”Musiken förändrar känsla och förväntningar totalt.”* Slutligen angavs att musiken upplevdes stå för merparten av den emotionella kommunikationen: *”Musiken är avgörande för känslostämning till säkert 90 % vilket märks tydligt i de olika klippen.”*

5. Diskussion

I detta kapitel diskuteras studiens resultat gällande emotionell precision vid användning av semantiska och strukturella promptingstrategier. Analysen belyser systemets tendens mot positiv valens och hur den audiovisuella kontexten påverkar deltagarnas upplevelse. Vidare granskas studiens metodologiska begränsningar samt AI-musikens roll som ett verktyg kontra de etiska riskerna för det konstnärliga hantverket. Avsnittet avslutas med förslag på framtida forskning.

5.1 Resultatdiskussion

5.1.1 Valens och arousal skattningar och emotionell precision

En central del i analysen är förhållandet mellan de två emotionella dimensionerna valens och arousal. Tidigare forskning på generativa AI-modeller, däribland Hu [11], har indikerat att systemen generellt uppvisar en högre förmåga att fånga arousal än valens. Studiens resultat för kategorin *Sorg* utgör ett tydligt stöd för denna tes. Som framgår av Tabell 4 (se 4.1) låg medelvärdena för arousal ($M = 3,03$ & $M = 3,55$) betydligt närmare målet ($A = 2,0$) än vad valensvärdena gjorde ($M = 4,58$ & $M = 5,06$). Här lyckades Suno tolka instruktioner för låg energi betydligt bättre för båda strategierna än den lyckades att fånga den negativa valensen.

Däremot nyanseras Hus slutsats av resultaten för *Lugn*. Här uppvisade den semantiska strategin en närmast perfekt precision på valensaxeln ($M = 5,94$) mot målet ($V = 6$) , medan det istället var en förhöjd arousal ($M = 3,90$) mot ($A = 2,0$) som drog ner den totala precisionen. En möjlig orsak till detta kan vara ett av de mönster som identifierades av Hu [11], där en betydande majoritet (63,11 %) av den AI-genererade musiken i undersökningen skattades ha en positiv valens. Denna tendens mot behagliga känslolägen kan förklara varför precisionen för *Lugn* blev högre i denna studie då generativa modeller har en tendens att skapa musik som upplevs som behaglig och välproducerad. När målet istället var *Sorg* kan denna tendens istället bidragit till att det varit svårt att generera ett negativt eller obehagligt uttryck. Vilket kan vara en möjlig förklaring till att musiken istället stannade kvar i ett behagligt register som deltagarna snarare tolkade som lugnt eller romantiskt. Detta stöds även av analysen av ”Övrigt”-svaren (se 4.3.1), där ord som "Romantiskt" och "Kärleksfullt" användes för att beskriva klippen för *Sorg*. Dessa positiva beskrivningar korrelerar med den låga precisionen på valensaxeln och förklarar varför en betydande andel deltagare valde kategoriseringen ”Lugn”.

Det går även att finna stöd i den förhöjda arousalskattningen för *Lugn* i ”Övrigt”-svaren, där deltagarna använde termer som ”Triumf” och ”Äventyrlig” för att beskriva klippet *Semantisk Lugn*.

5.1.2 Strategiernas effektivitet för komplexa emotioner

Ett av studiens mest framträdande fynd är att den semantiska strategin var signifikant mer precis för att förmedla sorg än den strukturella strategin ($p = 0,041$). Detta är intressant i förhållande till Fišer et al. [7], som noterat att mer detaljerade instruktioner (likt den strukturella strategin) kan öka den kognitiva ansträngningen hos lyssnaren. En möjlig förklaring till detta är att AI-modellen har en starkare koppling mellan emotionella etiketter som "sorrow" och de musikaliska dragen som krävs för sorg, snarare än att enbart förlita sig på teoretiska parametrar. Det är möjligt att för många teoretiska parametrar snarare gör att AI-modellen försöker få med alla delar och därmed kan tappa den övergripande känslan som skulle förmedlas.

Den strukturella strategins begränsning blev tydlig i kategoriseringen (Figur 6), där en stor majoritet felaktigt valde ordet lugn för det strukturella sorg-klippet. Detta tyder på att de parametriska instruktionerna resulterade i en musik som var för välproducerad och stabil för att få till en negativ valens.

Det är dock väsentligt att poängtera att även om den semantiska strategin i kategorin sorg presterade bättre än den strukturella, uppvisade kategorin sorg som helhet den lägsta emotionella precisionen i studien. Dessutom problematiseras resultatet genom att den strukturella strategin för *Sorg* uppvisade en något högre kategorisk matchning, samtidigt som den semantiska strategin resulterade i statistiskt signifikant högre emotionell precision.

5.1.3 Audiovisuell kongruens och visuell dominans

Att skattningarna för kategorin sorg genomgående uppvisade en dragning mot positiv valens kan inte enbart härledas till AI-modellen, utan måste även förstås utifrån samspelet med bilden. Enligt Cohens [6] *Congruence-Association Model* söker åskådare/lyssnaren efter strukturella matchningar mellan ljud och bild för att skapa en enhetlig mening. Resultaten visar att denna process fungerade mycket effektivt för kategorierna *Spänning* och *Lugn*, men sämre för kategorin *Sorg*. En möjlig förklaring till detta är att musiken fann en naturlig visuell koppling i tågscenens karaktär.

För *Spänning* uppnåddes en nästan perfekt kategorisering trots dess negativa valens. En möjlig förklaring är att spänningens höga arousal fann en direkt koppling i tågets fysiska rörelse framåt. Deltagarnas beskrivningar om att musiken skapade en känsla av att *"en bov har riggat bomber på spåret"* eller att det kändes som en *"actionscen"* illustrerar hur musiken fungerade som ett narrativt schema [2] som gav rörelsen en dramatisk innebörd.

Även *Lugn* visade en mycket hög precision, där den semantiska strategin nådde en bred matchning på 96,8 %. De kvalitativa svaren, med ord som *"hoppfullt"*, *"harmoniskt"* och *"vackert"*, bekräftar att deltagarna upplevde en positiv känsla som var i kongruens med den vackra vyn. Detta indikerar att den AI-genererade musiken var som mest effektiv när den förstärkte en redan existerande visuell dynamik såsom tågets färd genom ett vackert landskap, vilket naturligt rymmer både spänning (rörelse) och lugn (estetik).

I kategorin *Sorg* uppstod däremot en konflikt då det var svårare för musiken att få en visuell koppling till bilden, då inneboende tragiska element saknas. Även avsaknaden av en tydlig narrativ kontext försvårar för musiken att på egen hand förmedla en sorglig känsla. Istället för att acceptera en sorgsen ton som saknade visuell grund, omtolkade deltagarna musiken för att uppnå kongruens med bilden. Oftast valdes då istället *"Lugn"* men också fritextsvar som *"Romantiskt"*, *"Kärleksfullt"* eller med referenser till en *"Harry Potter-liknande atmosfär"*.

Resultaten antyder att den AI-genererade musiken kunde förstärka eller nyansera en känsla som redan finns i bilden, men hade svårt att på egen hand vända bildens emotionella tecken utan en visuell koppling i en narrativt begränsad kontext.

5.2 Metoddiskussion

Studien använde ett snöbollsurval som resulterade i 31 deltagare. Urvalet kan betraktas som relativt balanserat vad gäller kön då 14 män och 17 kvinnor deltog. Gällande ålder så finns det både yngre representanter då 20 personer var under 35 år, samt äldre då 9 personer var över 50 år. Denna åldersspridning minskar risken för att resultaten enbart speglar en specifik generations musikaliska preferenser, även om spridningen inte är helt jämn. Deltagarna var dock mer homogena i sin musikaliska bakgrund där en majoritet (21 av 31) saknade erfarenhet av musikproduktion eller teori. Detta innebär att studien främst mäter hur gemene man uppfattar AI-musik, snarare än hur experter inom området analyserar den. Det är möjligt att resultaten kring de strukturella parametrarna hade sett annorlunda ut om en större andel av deltagarna varit professionella musiker med en mer analytisk lyssningsvana. För att motverka detta hade ett mer stratifierat urval kunnat användas, exempelvis genom att ha två tydliga grupper, en med professionella musiker och en med individer utan musikalisk erfarenhet. Vid ett sådant urval hade resultatet kunnat jämföras för att identifiera om musikalisk expertis påverkar precisionen för de strukturella parametrarna.

Utöver urvalet utgör avsaknaden av djupgående kvalitativa data som förklarar deltagarnas skattningar en begränsning i studiens datainsamling. Om studien hade genomförts på nytt, skulle de kvantitativa enkätsvaren ha kompletterats med semistrukturerade intervjuer. Genom att låta deltagarna förklara sina tankar och känslor i realtid under lyssningsprocessen, skulle studien kunnat få en djupare insyn i hur de skattade musiken och varför.

Gällande studiens interna validitet hade en alternativ metod för att stärka resultaten varit att inkludera en kontrollgrupp som fick skatta musiken utan det visuella materialet . Om en kontrollgrupp exempelvis hade identifierat sorg i enbart den genererade musiken, skulle det ha bekräftat att AI-modellen faktiskt lyckats med sin uppgift, men att den visuella kontexten var för stark för att tillåta den känslan att nå fram. Att isolera de auditiva och visuella variablerna på detta sätt hade gett ett tydligare svar på vad som faktiskt genererade den emotionella upplevelsen.

En annan kritisk aspekt i metodiken rör individens roll som medskapare och beslutsfattare. Vid generering i Suno v4.5 skapas alltid två alternativa kompositioner baserat på samma prompt. I denna studie innebar det ett manuellt val av en av de två versionerna som bedömdes ha störst potential att förmedla den avsedda känslan. Därefter valdes även det mest representativa segmentet om 30 sekunder ut från den valda versionen. Detta introducerar en oundviklig mänsklig bias, då resultaten därmed inte enbart mäter AI-modellens råa förmåga, utan även den individuella förmågan att identifiera lyckat material. Det är möjligt att en annan forskare hade valt den alternativa kompositionen eller ett annat tidssegment. Detta understryker att den emotionella precisionen i denna studie är ett resultat av ett samspel mellan människa och AI som kan se annorlunda ut beroende på utövaren.

När det kommer till det använda visuella materialet kan valet av bild ha haft en starkare inverkan på resultatet än vad som initialt förutsågs. Även om bilden valdes för att vara neutral, visar diskussionen i 5.1.3 att den snarare kan ha fungerat som en dominant ram. Förstudien, som syftade till att identifiera en neutral visuell stimuli, begränsades av ett lågt deltagarantal, vilket medför att bildens neutralitet inte kan generaliseras med säkerhet. Ett betydligt större urval hade behövts för att statistiskt säkerställa att klippets värde för valens och arousal låg i mitten på studiens använda skala. Dessutom hade kriterierna för val av visuell stimuli kunnat utgå från etablerade databaser för validerat

bildmaterial, snarare än ett subjektivt urval, för att säkerställa att bilden inte påför en egen emotionell bias.

Slutligen formulerades samtliga prompts på engelska, då AI-modellen är tränad på engelskspråkig data. Samtidigt genomfördes enkäten och analysen på svenska. Även om termer som "sorg" och "sorrow" eller "lugn" och "calm" ligger nära varandra, kan nyanser ha gått förlorade i översättningen eller i deltagarnas tolkning av de svenska begreppen.

5.3 Studiens bidrag och reflektioner

Studiens resultat ger ett tydligt svar på forskningsfrågan genom att visa att *semantisk emotionell prompting* inte bara är likvärdig, utan i vissa fall (som vid sorg) är mer effektiv än en *strukturell parametrisk strategi*. Detta tyder på att det för en användare räcker att beskriva en övergripande känsla snarare än att besitta musikteoretiska kunskaper om enskilda parametrar för att styra AI-systemets emotionella uttryck.

I praktiken innebär detta fler möjligheter för skapare inom audiovisuell media. Genom att sänka tröskeln för att skapa emotionellt precis musik blir verktygen tillgängliga för en bredare grupp. Resultatet indikerar att AI-modeller som Suno v4.5 har blivit så pass bra på att tolka mänskliga känslolägen att de musikaliska byggstenarna som exempelvis tempo, tonart och instrumentation, hanteras effektivt av enkla semantiska ord.

Trots AI-systemets förmåga att tolka instruktioner visar studien att den mänskliga intentionen inte helt ersätts, utan snarare tar en annan form. Behovet av att först välja mellan olika kompositionsförslag och sedan extrahera specifika segment pekar på att det konstnärliga hantverket i en AI-kontext skiftar från komposition till kuration. I en audiovisuell kontext kan då AI:n föreslå musikaliska lösningar, men det krävs ett mänskligt öra och öga för att avgöra vilken av dessa som bäst uppnår kongruens med bildens narrativ. Detta ger stöd åt Anantrasirichai och Bulls [1] argument om att AI främst bör ses som ett assisterande verktyg som kräver mänsklig vägledning snarare än att bli en ersättare.

Trots att människan har en roll i skapandeprocessen är den på flera sätt annorlunda mot en kompositörs roll i traditionell mening. Detta väcker frågor kring vad det menas med att skapa musik och om användning av generativa AI-modeller faktiskt kan räknas som musikskapande. Inom traditionell komposition är processen hantverksmässig, där varje val är ett direkt uttryck för kompositörens intention. När interaktionen istället sker genom indirekta prompts, kan man argumentera för att användaren snarare agerar som en beställare av musik. Detta går att likna med hur en regissör beställer ett verk av en mänsklig kompositör men att den stora skillnaden är samspelet mellan regissör och kompositör, samt den personliga prägel som enskilda kompositörer besitter. Stora kompositörer som John Williams eller Hans Zimmer kännetecknas ofta av sin egna musikaliska prägel, som går att känna igen i deras verk oavsett genre. En stor förlust med att använda AI-genererad musik blir därmed att denna särskilda prägel riskerar att försvinna. Detta ligger i linje med den kritik som Holzapfel et al. [10] för fram gällande risken att det konstnärliga hantverket devalveras och blir mindre personlig.

Utöver rollproblematisering finns ett existentiellt perspektiv på vad som går förlorat när kreativa aktiviteter automatiseras. Att lära sig ett instrument eller bemästra musikteori är en krävande process, men det är också där en stor del av den mänskliga upplevelsen och stoltheten ligger. Att utvecklas inom ett musikaliskt hantverk kan dessutom vara ett viktigt sätt för en människa att uttrycka sina egna känslor. Det finns stora risker med att låta datorer ta över aktiviteter som anses vara lustfyllda och kreativa. När vi inte längre behöver anstränga oss för att skapa, riskerar vi att devalvera värdet av själva skapandet.

Samtidigt som användningen av AI-genererad musik väcker dessa etiska frågor, indikerar resultatet av studien en positiv potential för småskaliga kreatörer. För en spelutvecklare eller filmskapare som arbetar ensam med begränsade resurser innebär AI-generering en ökad tillgänglighet av det kreativa uttrycket. I dessa sammanhang är målet sällan att ersätta en professionell kompositör, utan snarare att möjliggöra produktion där resurser för mänsklig expertis helt saknas. Här styrker studiens resultat att även användare utan musikteoretisk bakgrund kan uppnå hög emotionell precision genom enkla semantiska instruktioner. Detta innebär att en kreatör enklare kan skräddarsy musiken efter sitt narrativ istället för att förlita sig på att använda befintlig upphovsrättsfri musik (stock music). AI-genererad musik skulle i detta sammanhang kunna fungera som en kreativ möjliggörare som fyller det gap som uppstår när teknisk och teoretisk kunskap saknas men en tydlig emotionell vision finns.

5.4 Framtida forskning

Utifrån studiens resultat och de begränsningar som identifierats, föreslås följande områden för framtida studier:

Resultaten för kategorin *Sorg* pekar på att det är svårt för den AI-genererade musiken att på egen hand framkalla en obehaglig/negativ känsla utan en visuell koppling. En möjlig vinkel för framtida studier är därför att undersöka betydelsen av narrativ kontext, exempelvis genom att ge deltagarna en kort bakgrundshistoria till klippet.

Eftersom utvecklingen av generativa AI-system sker i ett mycket högt tempo, utgör denna studie endast en ögonblicksbild av Suno v4.5:s förmågor. I takt med nya versioner vore det värdefullt att repetera studien för att exempelvis se om den identifierade tendensen mot positiv valens minskar i takt med att modellerna tränas på mer varierad data. Vidare föreslås jämförande studier mellan olika typer av generativa modeller som exempelvis Udio eller MusicLM för att undersöka om vissa modeller är bättre lämpade för specifika emotionella uttryck eller promptingstrategier.

Ur ett användarperspektiv vore det även relevant att studera iterativ prompting. Denna studie fokuserade på AI-musikens precision vid en enda prompt, men i verkligheten interagerar användaren ofta med systemet genom flera omgångar av instruktioner. Forskning kring hur mänsklig intention utvecklas och preciseras genom en dialog med AI:n skulle kunna bidra till kunskap om människans roll kan skifta mot en mer aktiv "medskapare", snarare än som en “beställare” som diskuteras i 5.3.

Slutligen föreslås forskning som undersöker hur musikalisk kvalitet påverkar den känslomässiga upplevelsen. Genom att låta deltagarna skatta kvalitén på musiken, kan man undersöka om hög ljudmässig kvalitet bidrar till en bättre emotionell precision. Det vore även intressant att studera om AI-systemets tendens att skapa musik som låter snyggt och välproducerat faktiskt kan vara ett hinder när man vill förmedla mer mörka eller obehagliga känslolägen.

Slutligen föreslås studier som undersöker kulturella skillnader i tolkningen av AI-genererad musik. Då nuvarande modeller främst är tränade på västerländsk musiktradition [10], vore det värdefullt att undersöka om den emotionella precisionen förblir densamma för lyssnare från andra musikaliska kulturer. Detta skulle belysa om AI-modeller riskerar att göra all musik mer ensidig genom att bara utgå från västerländska mönster, vilket i sin tur hotar den musikaliska mångfalden.

6. Slutsats

Syftet med denna studie har varit att undersöka hur olika textbaserade promptingstrategier påverkar den emotionella precisionen i AI-genererad musik. Genom ett experiment där 31 deltagare utvärderade orkestral musik skapad i Suno v4.5 till en specifik videosekvens, har semantisk och strukturell prompting jämförts. Denna studie tyder på att *semantisk emotionell prompting* är likvärdig

och i vissa fall även bättre jämfört med *strukturell parametrisk prompting* för att styra emotionell precision i AI-genererad musik i en audiovisuell kontext. Resultaten indikerar att generativa modeller som Suno v4.5 har en god förmåga att tolka övergripande känslomässiga etiketter, vilket gör tekniken tillgänglig även för skapare som saknar musikteoretisk expertis.

Ett centralt fynd är att den semantiska strategin uppvisar högre precision vid komplexa känslor som sorg, medan båda strategierna var relativt likvärdiga för lugn och spänning. Studien belyser dock en inneboende utmaning i systemets tendens mot positiv valens och musik som låter välproducerad, vilket kan försvåra förmedlingen av negativa eller obehagliga känslouttryck. Vidare understryks betydelsen av den audiovisuella kontexten då resultatet tyder på att den genererade musiken i detta sammanhang fungerade som mest effektivt när den förstärker en befintlig visuell dynamik.

Teoretiskt bidrar studien med en nyanserad bild av förhållandet mellan valens och arousal i AI-musik. Praktiskt pekar resultaten på fler möjligheter för skapare inom audiovisuell media, där tröskeln för att skapa emotionellt precis musik sänks. Slutsatsen är att AI-musik kan vara ett kraftfullt verktyg för exempelvis småskalig medieproduktion, men att människans roll i musikskapandeprocessen förändras avsevärt. Även om människan inte helt står utanför processen, skiljer sig denna nya roll stort mot den traditionella rollen som kompositör. Detta väcker existentiella frågor om det konstnärliga skapandets värde, då den hantverksmässiga ansträngningen istället ersätts av algoritmiska förslag. Men också frågor kring avsaknad av den personliga konstnärliga prägeln, samt själva definitionen av vad det innebär att skapa musik.

Referenslista

[1] Nantheera Anantrasirichai and David Bull. 2022. Artificial intelligence in the creative industries: a review. *Artif. Intell. Rev.* 55, 1 (January 2022), 589–656. https://doi.org/10.1007/s10462-021-10039-7

[2] Marilyn G. Boltz. 2001. Musical Soundtracks as a Schematic Influence on the Cognitive Processing of Filmed Events. *Music Percept.* 18, 4 (June 2001), 427–454. https://doi.org/10.1525/mp.2001.18.4.427

[3] Roberto Bresin and Anders Friberg. 2011. Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex* 47, 9 (October 2011), 1068–1081. https://doi.org/10.1016/j.cortex.2011.05.009

[4] Blake Brittain. 2024. Music labels sue AI companies Suno, Udio for US copyright infringement. *Reuters*. Retrieved June 2, 2026 from https://www.reuters.com/technology/artificial-intelligence/music-labels-sue-ai-companies-suno-udio-us-copyright-infringement-2024-06-24/

[5] Michel Chion, Claudia Gorbman, and Walter Murch. 2019. *Audio-vision: sound on screen* (Second edition ed.). Columbia University Press, New York.

[6] Annabel J. Cohen. 2001. Music as a source of emotion in film. In *Music and emotion: Theory and research*, Patrik N. Juslin and John A. Sloboda (eds.). Oxford University Press, Oxford, 249–272.

[7] Nikolaj Fišer, Miguel Ángel Martín-Pascual, and Celia Andreu-Sánchez. 2025. Emotional impact of AI-generated vs. human-composed music in audiovisual media: A biometric and self-report study. *PLOS One* 20, 6 (June 2025), e0326498. https://doi.org/10.1371/journal.pone.0326498

[8] Thomas Fritz, Sebastian Jentschke, Nathalie Gosselin, Daniela Sammler, Isabelle Peretz, Robert Turner, Angela D. Friederici, and Stefan Koelsch. 2009. Universal Recognition of Three Basic Emotions in Music. *Curr. Biol.* 19, 7 (April 2009), 573–576. https://doi.org/10.1016/j.cub.2009.02.058

[9] Alf Gabrielsson and Erik Lindström. 2010. The role of structure in the musical expression of emotions. In *Handbook of music and emotion: Theory, research, applications*, Patrik N. Juslin and John A. Sloboda (eds.). Oxford University Press, Oxford, 367–400.

[10] Andre Holzapfel, Bob L. Sturm, and Mark Coeckelbergh. 2018. Ethical Dimensions of Music Information Retrieval Technology. *Trans. Int. Soc. Music Inf. Retr.* 1, 1 (September 2018), 44–55. https://doi.org/10.5334/tismir.13

[11] Xinyue Hu. 2025. Feels in the Music Machine: Exploring the Relationship between Text Prompts and AI-Generated Music in Terms of Valence and Arousal. Master's thesis. KTH Royal Institute of Technology, Stockholm. Retrieved from https://www.diva-portal.org

[12] Patrik N. Juslin and Daniel Västfjäll. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 5 (October 2008), 559–575. https://doi.org/10.1017/S0140525X08005293

[13] Kathryn Marie Kalinak. 2014. *Settling the Score: Music and the Classical Hollywood Film*. University of Wisconsin Press, Madison.

[14] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 21, 2020. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376739

[15] Donald A. Norman. 2013. *The design of everyday things* (Revised and expanded ed ed.). Basic Books, New York.

[16] James A. Russell. 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 6 (December 1980), 1161–1178. https://doi.org/10.1037/h0077714

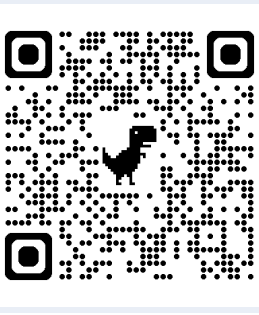
[17] Oliver Vitouch. 2001. When Your Ear Sets the Stage: Musical Context Effects in Film Perception. *Psychol. Music* 29, 1 (April 2001), 70–83. https://doi.org/10.1177/0305735601291005

Bilagor

Bilaga 1: Ljud- och videomaterial

I denna bilaga återfinns QR-koder till samtliga 6 klipp som använts i studien. Scanna QR-koden med din mobilkamera för att öppna klippet på YouTube.

Klipp	QR-kod & Länk
Semantisk Lugn	https://youtu.be/uBTrES_l2O8
Strukturell Lugn	https://youtu.be/X0icO7XYIbw
Semantisk Sorg	https://youtu.be/ComhtG-PIDw

Strukturell Sorg	 https://youtu.be/E4HqFswqxVs
Semantisk Spänning	 https://youtu.be/rJMNShtjx2g
Strukturell Spänning	 https://youtu.be/BGoNRK2_Urw

TRITA-EECS-EX-2026:395
Stockholm, Sverige 2026

www.kth.se