# Interpreting Nonverbal Cues Through Sound: Exploring Accessibility for BLV Individuals in Online Communication Using Cinematic Media and Sound Design

By Shamayma Mobin

Submitted to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Inclusive Design Toronto, Ontario, Canada, 2025

# Creative Commons Copyright notice

# Abstract

Nonverbal cues play a critical role in everyday communication, yet much of this information remains inaccessible to blind and low-vision (BLV) individuals, particularly in digital and remote settings. The study explores how alternative auditory strategies including tone of voice, sound effects, music, and verbal descriptions can potentially support the interpretation of nonverbal information in online communication platforms. The study investigates how these modes function in relation to key dimensions such as iconicity, affordance, multimodality, and embodied meaning-making. To examine perception, media clips from films and animated shows were used to identify how nonverbal cues are interpreted through characters' vocal tone, body movement, and facial expressions, and to analyze how integrated audio elements such as music and sound effects contribute or disrupt meaning. Through co-designing, these examples helped participants reflect on their own sense-making processes and highlight which cues were perceptible or missed. To transition to more real-life contexts, scripted Zoom call scenarios were introduced to simulate online communication, allowing participants to experience and evaluate how auditory cues adapted from their media interpretations could function in relatable, remote conversations. Findings reveal that each auditory alternative contributes differently to meaning-making: tone of voice offers expressive immediacy, sound effects can convey action or emotional tone (with varying levels of abstraction), music supports mood-setting and affective interpretation, and verbal descriptions help fill semantic and contextual gaps. Tables composed of summarised findings and future recommendations were developed for a deeper understanding of the affordances and limitations of the sound cues in nonverbal cue accessibility for BLV individuals and also to serve as a foundation for designing more inclusive communication systems in online communication spaces.

# Acknowledgements

**Allah Almighty**
My greatest and infinite gratitude to Allah Almighty, for blessing me with the strength and patience to embark on this journey. For blessing me the opportunity and guidance in helping me accomplish my goals.

**Family**
To my family, I am deeply grateful for your love and support. For motivating me to pursue this Master's program. I would not have been able to complete my Master's without you.

**Dr. Peter Coppin, Primary Advisor**
Thank you Peter, for your invaluable guidance and support throughout this project. Your thoughtful insights and feedback were instrumental in shaping the direction of my research. Working under your supervision was a great learning experience that allowed me to deeply engage with and explore my topic of interest in a meaningful and focused way.

**David Barter, Teaching Assistant III & IV (Tutorial Leader)**
David, you were more of a secondary advisor for this research. Your contributions were deeply appreciated and impactful. Thank you for always being readily available to help, whether it was through feedback, thoughtful suggestions helped me refine my ideas and strengthened the overall direction of the project. I am sincerely grateful for your support throughout this process.

**Perceptual Artifacts Lab (PAL) Research Group, OCAD University (led by Dr. Peter Coppin)**
Thank you to the PAL research group, these weekly research meetings were a great space for new learnings and getting insightful feedback from a dedicated group of interdisciplinary researchers.

**Participants**
I would also like to extend my heartfelt thanks to all the participants who took part in this research. Your willingness to share your time and insightful experiences made this project not only possible but deeply meaningful. I am truly grateful for your openness and generosity in contributing to a project aimed at improving accessibility and communication for the blind and low-vision community.

**A Study of Accessible and Inclusive Virtual and Blended Information and Communication Technologies (ICTs) for the Federal Public Service and Federally Regulated Industries in Post-COVID-19 Canada (ASC Grant #102194)**
This research builds upon the work from that grant project that was conducted prior to your Masters degree

# Table of Contents

# List of Figures



# List of Tables

# 1.0 Introduction

## 1.1 Project background and purpose

Nonverbal communication is essential for understanding social interactions, emotions, and intent. It includes a range of cues: facial expressions, gestures, posture, eye contact (visual cues), tone of voice, silence (auditory cues), and touch, spatial cues (tactile and spatial cues) that supplement and sometimes even replace verbal language (Knapp, Hall & Horgan, 2013). These cues may also be considered as spatial-topological properties (S-Tl; Lee, Sukhai and Coppin, 2022) because they are properties of people, objects and environments that are occupying space and the relations between these that are intentionally shaped by human behaviours and thus convey their intentions. However, for blind and low vision (BLV) individuals, interpretation gaps arise when access to S-T properties such as visual cues is constrained, however it becomes more challenging when alternative and compensatory cues such as auditory and haptic cues (Qiu, 2019) also become disrupted. This is observable in ICT platforms in which spatial-topological (S-T) properties such as auditory proxemic cues (personal physical space and with others) are absent due to restricted interface facilities, which poses a significant challenge in understanding interactions (Lee, Sukhai and Coppin, 2022). Online environments such as video calls often limit the availability or clarity of auditory nonverbal cues, caused by technical issues such as poor audio quality which leads to communication issues such as lack of engagement, and focus which results in video conferencing being less productive than face-to-face meetings (Luebstorf, 2023).

In-person settings allow BLV individuals to make use of proximate auditory information - paralinguistic cues such as tone, rhythm, pace, and vocal inflections which also infer other nonverbal cues, such as head orientation ,and serve as compensatory modalities (Qiu, 2020). Paralinguistics refers to the vocal elements accompanying speech that convey emotional nuance and intention (Mamurova, 2024). The affordances that tone of voice in speech offers (e.g., a rise in tone indicating surprise or enthusiasm) may be disrupted or diminished due to digital filtering and lack of spatial context in ICT platforms (Dash, 2022; Lee, Sukhai and Coppin, 2022). Therefore, this project explores how BLV individuals perceive nonverbal cues in online communication platforms, in which visual modalities are often inaccessible. With growing reliance on video conferencing and online communication, the social gap created by limited access to nonverbal information such as facial expressions, gestures, and visual context has become more pronounced.

To investigate and understand how nonverbal cues can be perceived and interpreted auditorily, media, particularly narrative video content such as films, television series, and online videos, were used as a medium for observation. Media such as films mirrors real-world communication (Chion, 1994) while in other examples e.g., cartoons, it is often dramatized but rich in layers of nonverbal expression (Sagheer, 2024). It provides controlled, replayable scenarios where conversations unfold with accompanying gestures, facial expressions, tone changes, music, and sound effects (Chion, 1994). These media experiences

can offer a space to observe their interpretation of cues and identify what information can or cannot be accessed through sound alone. Unlike live, unpredictable environments, media or pre-recorded videos offers consistency and intentional design, making it easier to isolate and analyze specific nonverbal elements and their affordances (Birdwhistell, 1952).

The study also investigates how various auditory elements, including tone of voice, sound effects, and music, contribute to or hinder interpretation of these cues. Using entertainment media for observations and scripted Zoom calls as a transitional setting, this research identifies both the challenges and opportunities in enhancing accessibility. The progression from media observation to structured co-design testing supports the development of novel audio-based accessibility strategies that are both grounded in real experience and informed by experimental reflection. The relevance of this study lies in its contribution to inclusive communication design, particularly within ICT platforms, by examining alternative auditory strategies that support social understanding for BLV individuals.

The investigation revealed that BLV participants primarily rely on paralinguistic vocal cues such as tone, rhythm, and pauses to interpret emotion and intent in conversations. This was confirmed from the media examples, in which participants missed crucial visual gestures and facial expressions but could often perceive character emotion and interaction dynamics through tone of voice; however, when perception issues arose from tone of voice, participants received varying levels of feedback from certain sound effects, music and descriptive support. Scripted Zoom call experiments showed that while tone and vocal rhythm conveyed some affective information, integrated auditory enhancements (e.g., sound effects for gestures or descriptions) enhanced feelings and certainty of interpretation. Participants responded positively to the co-designed cues, particularly those that added nuance to ambiguous or emotionally charged moments, suggesting a strong potential for improving accessibility in online communication environments.

## 1.2 Context

This section provides an overview that identifies the challenges that BLV individuals face when accessing visual nonverbal cues and problems that arise when doing so in online meetings. This sets the problem space in which this project attempts to make an impact.

A 2017 report by the World Health Organization (WHO) estimated that there were 253 million visually impaired individuals worldwide, with 36 million being blind and 217 million having low vision (Qiu, 2019). Among the challenges faced by BLV people, a significant one is their difficulty in perceiving nonverbal cues during face-to-face interactions. Nonverbal cues, which convey attitudes and interactions, heavily rely on visual cues such as eye movements, facial expressions, gestures, and body language, making them inaccessible to blind individuals and challenging for those with low vision (Vinciarelli, 2009). This limitation can lead to inconvenience in participating in conversations and may result in impatience or misunderstanding from sighted counterparts (Baumeister, 1995). Consequently, many BLV individuals resort to passive communication strategies, in which they are only

listening during conversations (Goharrizi, 2010). According to Griffin's Uncertainty Reduction theory, this suggests that the absence of visual cues can lead to uncertainty about the attitudes of sighted individuals, potentially causing communication breakdowns and lowering self-confidence (Humphrey, 2015). Additionally, studies have observed differences in behavior between BLV and sighted individuals during face-to-face interactions, with BLV individuals often exhibiting more introverted and submissive tendencies (Qiu, 2019; Kemp, 1986).

After recognizing the fundamental human need for social connection, assistive systems for BLV individuals have naturally focused more on fulfilling basic needs such as navigation and access to information (Brock, 2013; Galioto, 2018; Botzer, 2018). While some studies have explored technology-based solutions to assist BLV individuals, there remains a lack of comprehensive empirical understanding of their real-life social needs, including both limitations and capabilities (Qiu, 2020), especially in online platforms. Ever since the COVID-19 pandemic that began in March 2020, the shift to online meetings has made a significant impact on our lives, not only because it was sudden (Reed, 2022) but also because it was necessary (Anderson, 2020). In real-world interactions, spatial and topological (S-T) cues such as body posture, facial expressions, environmental context, and object arrangement play a crucial role in shaping our understanding of others' emotions and intentions (Lee, Sukhai and Coppin, 2022). The study shares an example of a person simply entering a manager’s office allows one to gauge stress levels or approachability through nonverbal and environmental cues. However, in virtual settings, these rich perceptual signals are often lost or poorly translated through video conferencing tools, which restrict access to such contextual information. As a result, users in online meetings rely heavily on limited visual or verbal data, impairing their ability to interpret nuanced interpersonal dynamics and intentions effectively. This applies to every individual, regardless of their ability and disability; however, in the specific case of those with disabilities, online platforms, particularly for meeting platforms such as Zoom, can be more enhanced for better social interaction experiences (Kim and Taylor, 2024).

The reason for this is that misunderstandings and interpersonal complications arise as these nonverbal signals are increasingly missed in online settings — and not only because our mediums lack fidelity (Park & Whiting, 2020). This is more problematic for BLV individuals who mostly depend on one sensory channel they can rely on – auditory. A study by (Qiu, 2019) confirms how BLV individuals rely more on audible cues and thus the lack of auditory fidelity in ICTs holds greater consequence for them. Consequently, visual conversation cues (VCCs) in video calls are not accessible to BLV individuals, which can lead to conversational asymmetry, particularly with sighted people. (Park & Whiting, 2020).

Therefore, this gap in perceptual access to nonverbal cues or spatial and topological properties (S-T) such as gestures, facial expressions and tone for BLV individuals forms the foundational problem addressed in this project. In virtual meetings, where interfaces constrain what can be perceived, the absence of these (S-T) properties becomes more critical for BLV users, who already rely on alternative sensory input (Lee, Sukhai and Coppin, 2022).

While considerable research has explored how blind and low-vision (BLV) individuals access nonverbal cues, there remains a notable gap in studies specifically examining their experiences in accessing nonverbal cues within online meeting environments. Moreover, concrete measures to address these accessibility challenges in digital contexts are still limited. It is for this reason that, before ideating and developing any assistive technologies or bettering existing online platforms, it is crucial to first understand these insights as they allow the designer, developer to empathize with BLV users, contextualize design challenges, and identify new opportunities for assistive technology. Without this understanding, assistive devices may inadvertently emphasize users' disabilities, potentially leading to negative social experiences (Shinohara, 2011). Therefore, it is essential to gain a nuanced understanding of BLV individuals' lived experiences concerning nonverbal cue accessibility in online communication platforms, considering both their limitations and capabilities in real-life contexts.

## 1.3 Objectives

The study seeks to understand the capabilities and difficulties in accessing nonverbal cues in ICT platforms and explore what multimodal means can be used to address any gaps emerging from interpretive challenges. The project is guided by the following objectives:

1. To examine which visual and auditory nonverbal cues are most effectively interpreted or often missed across three communicative contexts: media representations, in-person interactions, and online video calls.
2. To evaluate the effectiveness of additional integrated audio modalities—such as sound effects, tone of voice, music, or descriptions—in aiding the interpretation of nonverbal cues.
3. To develop informed suggestions and recommendations for enhancing the clarity and certainty of nonverbal cue interpretation for BLV individuals during online and mediated conversations.

## 1.4 A preview of Findings

Table 4 highlights the affordance potential implications in ICT platforms.

| | | Findings | Implications |
|---|---|---|---|
| **Descriptions** | Literal, concise | Literal concise descriptions were supported because they provided space for independently interpreting the meaning of the cues.<br>They were direct indicators linguistically describing expressed visual cues<br>Risk of redundant information is less | **"She's nodding"** — conveys agreement or understanding.<br>**"He's raising his eyebrows"** — indicates surprise or curiosity.<br>**"They're clapping"** — signals approval or celebration. |
| | Detailed with affordances | Description of nonverbal visual cue and potential affordances can be effective for blind people who are conceptually unfamiliar with the associated meaning of the cues (when interpretation from literal descriptions remains ambiguous, incomplete).<br>Acts as a bridge for informing meaning or confirming interpreted meaning of described cues | **She's slowly nodding with her arms crossed, signifying she is getting angry"** — may convey reluctant agreement.<br>**"He throws his hands in the air and rolls his eyes, signifying frustration"** — expresses frustration or sarcasm.<br>**"Her smile fades and she turns away slightly meaning she is sad"** — suggests discomfort or disagreement. |
| **Sound effects** | Concrete | Perceptually immediate and specific.<br>They afford instant recognition and are effective for direct indexical interpretation, identifying clear physical actions non-linguistically.<br>Easier to learn and becoming familiar with when used for signifying actions. | **"Clapping sound"** — signifying someone clapping.<br>**"Crushing sound"** — someone gesturing crushing something.<br>**"Party music sounds"** — someone gesturing upbeat dance moves |
| | Abstract | Abstract sounds are only effective when one becomes familiar with the meanings associated with it. When the abstract sounds are repeated to an extent that the perceiver can instantly perceive it and understand the meaning,<br>Can be effective for non-linguistically conveying visually expressed reactions in online meetings; a person's questioning look, combined shock reaction, silently clapping | **"Ding"** — a participant has a realisation or an idea.<br>**"Whoosh"** — someone gestured quickly or shifted position dramatically.<br>**"Pop"** — someone have an questioning look |
| **Music** | | More conceptually specific in understanding.<br>Effective for setting a mood, enhancing feelings.<br>Becoming familiar with its affordance so it can be instantly interpreted. | **A short upbeat jingle** — signals a cheerful or successful moment, like a task being completed or a warm welcome.<br>**Slow, melancholic string tone** — conveys disappointment, reflection, or tension.<br>**Mysterious low-pitched hum** — signals confusion or an unresolved issue in the conversation. |

Table 1: Suggestive Implications for Future Work

The Table emphasizes which cues are most effective in conveying emotional (e.g., feelings, moods) and situational (e.g., actions, gestures, and indexical meanings) information. Notably, the study uncovered a new and meaningful insight: the *affective affordance* of audio cues. This refers to their ability to communicate the emotional impact or intention of a nonverbal cue, even when the specific gestures or facial expressions are not explicitly identified.

The findings reveal that tone of voice serves as an alternative and compensatory cue for interpreting information which is also expressed from visual cues. However, in the absence of tone of voice or disruption, descriptions, sound effects and music are relied on as an auditory compensatory cue. Linguistic descriptions allow space for personal interpretation of visual cues, but can lead to ambiguity when the listener lacks prior conceptual familiarity. In such cases, more detailed descriptions help clarify meaning. Different use of sound effects afford different uses; Concrete sound effects (sound used with its origin source) support recognition of physical actions (indexical cues), while abstract sound effects (sound used on a non origin) contribute to emotional or atmospheric understanding (affective affordances), though both require familiarity for accurate interpretation. Music was also found to play a meaningful role in conveying emotional tone and supplying emotional context.

## 1.5 Significance

The significance of this research lies in its contribution to understanding and improving the accessibility of nonverbal communication for blind and low vision (BLV) individuals in online communication platforms. This study fills that gap by investigating the interpretation of nonverbal cues across media, identifying which cues are missed or misinterpreted, and evaluating the potential of supplementary audio strategies such as sound effects, music, and descriptive language—to enhance comprehension. The findings inform not only accessibility

design for online communication tools but also contribute to broader discussions in social semiotics, auditory perception, and inclusive design. Ultimately, this research advocates for a multimodal approach to digital interaction, aiming to reduce communicative inequities and support more socially inclusive virtual environments for BLV individuals.

The study not only presented the effectiveness of using cinematic media in helping participants recall past communication experiences, but also enabled them to identify specific visual cues they typically miss—details that might have been difficult to access without this method. Importantly, the findings highlight the powerful role of tone of voice in conveying meaning. Participants also reflected on when and how audio cues can enhance communication in online settings, particularly during interactions with unfamiliar individuals. Insights from the co-design sessions further emphasized the potential of integrating certain audio cues such as sound effects not only to directly represent the intended meaning of visual or paralinguistic expressions, but also to provide enough context for BLV users to interpret the cues autonomously. These findings point to promising strategies for making online meetings more accessible, inclusive, and socially intuitive for BLV individuals.

Table 4, outlining various audio cues ranging from sound effects to spoken descriptions and their conceptual, perceptual affordances as well as possible implications offer a valuable foundation for accessibility design in online meetings. By categorizing how each cue functions in conveying emotional and situational meaning, the table serves as a practical reference for identifying which sounds are most effective in supplementing or replacing visual information. Notably, it introduces the concept of *affective affordance*, showing that some audio cues can communicate emotion or intention even when the exact gesture or facial expression is not discernible. This resource can guide developers, designers, and accessibility practitioners in crafting more intuitive and inclusive communication tools for BLV individuals, marking a concrete step toward improving social participation and autonomy in online platforms.

## 1.6 Limitations and scope

The limitations of this study resulted from scoping a project that was feasible to accomplish within the Master's major research project timeline. Limitations included the theoretical scope and scope of the literature drawn upon.

This study is bounded by its qualitative nature and limited participant sample, focusing on the interpretive experiences of individuals rather than producing generalizable data – the data showed mixed findings therefore, a greater number of sessions needs to be done. In addition to this, the findings are exploratory and context-specific because of the reliance on curated media examples rather than live meeting data. Furthermore, while the project proposes potential audio interventions, it does not include the development or testing of real-time technological implementations. However, these boundaries were intentional to prioritize depth of understanding over breadth at this stage of investigation.

The scope of the literature drawn for this study was limited to what supported the method of using cinematic media and pre-recorded videos for studying nonverbal perception and interpretation through multimodal audio cues, as well as supported the observations from the conducted research activities. This MRP did not entirely draw upon the extensive histories and breadth of work on shared nonverbal cues perception, accessibility interventions in ICT platforms and cognitive semiotics understanding.

## 1.7 Outline of the Study

Section 1.0 introduces the study by outlining the accessibility challenges BLV individuals face in interpreting nonverbal cues, particularly in online meetings, and highlights the limited existing literature documenting their lived experiences in these contexts. Section 2.0 introduces the past practices that provide the foundation for the hypothesis of using cinematic media as a method for studying BLV people's perception and interpretation of nonverbal cues and if sound design can be an effective alternative medium for accessing visually expressed information. Following this, in section 3.0, an overview of the methodology and research activities will be illustrated which includes details about interview sessions and co-design sessions. The section will also present co-design sessions comprising two phases where different media and other pre-recorded video examples will be used for studying. After this, section 4.0 will set out the findings and section 5.0 will involve a discussion of those findings and will be presented in tables. Lastly, conclusion of the study and recommendations for the future will be discussed in sections 6.0 and 7.0.

# 2.0 Theories and key concepts from the literature

This section outlines the motivation of using cinematic media as an alternate medium for studying nonverbal cues perception with the participants. The section supports how pre-recorded videos can be used as an effective medium for observing accessibility and interpretation. It also reveals how past research has studied sound design in cinema films as an effective mode for conveying situational, emotional information when anchored with visuals and also independently. As a result, a theory emerged based on literature review if sound design like sound effects and other audio cues like descriptions can serve as a potential alternative for conveying information when missed due to lack of perceptual access to visual cues.

## 2.1 Media for Studying Nonverbal Cues

Since this research is qualitative, the goal of the researcher is to attempt to access the thoughts and feelings of the participants. This becomes a challenging task, as it involves asking people to talk about things that may be very personal to them. Sometimes the experiences being explored are fresh in the participant's mind, whereas on other occasions reliving past experiences may be difficult. (Sutton & Austin, 2015). Conversing about

capabilities and limitations of perceiving nonverbal cues and comparing the accessing experience in face-to-face and online settings can be precarious for some participants since most of these experiences happen daily. However, since the primary source of data collection is from participants' past experiences and due to difficulties in recalling, the research utilises an alternate approach for not only prompting the participants in recalling past experiences but also to observe immediate interpretations – through media films.

The practice of studying nonverbal cues from pre-recorded videos became famous from one of the works done by Birdwhistell (1952). In his seminal work, Birdwhistell pioneered the systematic study of nonverbal communication, which he termed "kinesics." Recognizing the limitations of real-time observation, Birdwhistell employed film as a critical research tool to capture and analyse the nuances of body movements and gestures in interpersonal interactions. By recording social interactions on film, he could meticulously review and annotate subtle nonverbal cues such as posture shifts, facial expressions, and hand gestures that might otherwise go unnoticed. This method allowed for repeated analysis, enabling researchers to identify patterns and contextual meanings within nonverbal behaviour. Birdwhistell's innovative use of film not only enhanced the accuracy of behavioural analysis but also laid the groundwork for future studies in visual anthropology and communication, emphasizing the importance of visual media in understanding human interaction.

In another example, Caldwell (2022) compiled a series of studies that apply Systemic Functional Linguistics (SFL) to various real-world contexts, including the nuanced analysis of nonverbal communication. The volume emphasizes the importance of multimodal discourse analysis, recognizing that meaning is constructed not only through language but also through other semiotic resources such as gesture, posture, facial expressions, and visual framing. One of the studies in the volume uses a fifteen-minute YouTube video of former Australian Prime Minister Julia Gillard's 'misogyny speech' (Gillard, 2012) as its dataset to illustrate the analytical tools presented. This example demonstrates how nonverbal cues such as gestures, facial expressions, gaze, and vocal tone interplay with spoken language to convey emphasis, emotion, and stance. Through the integration of video media as both source and subject, the researchers dissect how various semiotic modes contribute to meaning making in political discourse, showing the potential of media as a natural corpus for studying real-life nonverbal behaviour. This reinforces the book's core argument that communication is inherently multimodal and context-dependent, and that media provides a rich platform for capturing this complexity.

In the study done by (Naufaldi, 2022), the researchers employed the film Harry Potter and the Sorcerer's Stone as a medium to explore the intricate relationship between verbal and nonverbal communication. Utilizing a qualitative approach grounded in multimodal discourse analysis, the researchers analyzed interactions among characters to discern how meaning is co-constructed through spoken language and nonverbal cues such as gestures, facial expressions, and tone of voice. Their findings revealed that nonverbal elements significantly influence the interpretation of verbal messages, affecting aspects like politeness, implicit meaning, and the transformation of sentence forms. This study underscores the efficacy of

using film as a naturalistic setting to examine the dynamic interplay between verbal and nonverbal communication, providing valuable insights into how meaning is negotiated in real-life interactions.

## 2.2 Cinematic Films and Perceptual Experiences

Cinematic media refers to any medium that uses visual storytelling techniques typically associated with filmmaking, creating a movie-like experience. This can include films, television shows, commercials, music videos, and even certain forms of digital media. Cinema realists focus on films' strong tie to reality because of (various aspects of) its visual and aural presentation of information. They propose that films can get at—or show—reality in a way that other art forms cannot. The strongest versions of cinematic realism prioritize physical reality by making the bold claim that by virtue of the mechanical, photographic process of their creation, films put the viewer in a perceptual contact with things in the world (Fiorelli, 2016).

Film-viewing is a unique aesthetic experience. In a movie scene, media that are associated with other art forms individually—sound, language, images, narrative— act together, with their distinctive varieties of meanings informing one another and conveying natural meaning (Fiorelli, 2016; Grice, 1957). Natural meaning is defined as the reliable correlation between signs and what they signify, which is crucial for understanding film's realism. Natural meaning exists where one thing or property is a reliable indicator or sign of something else: stable correlations between information states allow one state to show something about another. We observe natural meaning everywhere, for instance, in facial expressions (smiles mean happiness, frowns mean sadness) and in things that stem from more culturally specific, or conventional, relations (Grice, 1957). Filmic natural meaning, then, is pervasive: as viewers, we detect it perceptually—by deploying our everyday recognitional processes. Insofar as natural meaning exists in the world— regardless of our desires or aims—it is tied to reality. And in turn, film is realistic because it pervasively trades on and presents us with instances of natural relations among information states. We do all of this via everyday perceptual processes, and this can all be attributed in general to the fact that films largely present the same basic sorts of cues as we encounter in real life—sights and sounds of objects and their properties. We gather information, including natural meaning, by seeing (e.g., people, their reactions and actions) and hearing (laughing, crying, running, clapping). Being able to identify objects and their properties (and events) in images is the same capacity at work in our ordinary perceptual experiences (Fiorelli, 2016).

Many theorists attempt to perceive films with different focuses; with cinematic realists focusing on film’s perceptual content, semioticians focusing on how movies communicate, and narrative theorists focusing on how we cognize a film’s fiction, and each of them engaging in those analyses independent of the others. While they vary in highly divergent ways, all versions share one common aspect: they prioritize film’s perceptual nature, pointing to the sights and sounds movies actually show us (Fiorelli et al. 2016).

## 2.3 Sound Design

Sound made its dramatic entrance in cinema in 1927, marking a significant shift in film production and perception. Chion (2001) explores how sound design functions not merely as a supplement to visuals but as a powerful conveyor of information in its own right. He introduces the concept of "added value," where sound infuses images with meaning that they would not carry alone. Chion (2001) emphasizes that sound design, including ambient noise, sound effects, voice, and music guides the viewer's perception, attention, and emotional interpretation. For instance, he discusses how off-screen sounds can suggest unseen actions or spaces, thus expanding the narrative beyond the visual frame (Chion, 2001). Sound also cues the viewer into emotional subtext, character states, or spatial orientation, often more effectively than visuals can on their own. This auditory dimension can operate independently, supporting or contradicting what is seen on screen, thereby enriching the interpretive possibilities for audiences (Chion, 2001). Chion's (2001) analysis positions sound as an active agent in meaning-making, crucial to the viewer's experience and understanding of cinematic moments.

Due to sound's affordances, many works have been developed using purely sound. A study conducted by Lopez (2022) explores how sound design can be used for providing accessible versions of films (Pearl) for BLV audiences as an alternative to Audio Description (AD) practices. In another study (Lopez et al., 2009), an audio film was developed with the aim of eliminating the need of visual elements and of a describer, by providing information solely through sound, sound processing and spatialization, and which might be considered as an alternative to Audio Description. Two reasons for this are the affordance of sound and limitations of Audio Descriptions. In the case of sound, the study was motivated by media such as Radio Drama and Audio Games which employ sounds from surrounding to convey information (Lopez & Pauletto, 2009a, 2009b, 2010; Lopez, 2015). With the exception of some visual elements present in audio games, sound is the main means of communicating the storyline and aiding gameplay (Drossos, 2015). These examples are effective because the experience they provide is designed to be accessible from the onset and goes to show how the fields of audio films and audio games consider sound design as an accessibility method that is integral to the creative process. (Lopez, 2022).

In conclusion, these studies show how pre-recorded media can be an effective tool for studying nonverbal cues in depth and how films, which possess a certain resemblance to real world perceptual experiences can be used as a platform for observing nonverbal cue accessibility. The sound design element also provides another reason for using films for observing the effectiveness of different types of audio cues for conveying information.

# 3.0 Methodology

The previous section outlined the theories and concepts that lay the foundation for the table developed in this study. This section will now present the research methods through which the previously identified concepts were processed and synthesized for the final development of the table..

## 3.1 Study design

Figure 1, Study Design Methodology, presents the study process. Open-ended interviews and Semi-structured Interviews were conducted in Step 1 for identifying the problem and understanding the challenges from the experiences shared by participants. For Step 2, Co-design was employed as a participatory method to collaboratively explore how BLV individuals perceive and interpret nonverbal cues as well as through audio. Its effectiveness lies in centring users' lived experiences, enabling contextually grounded and accessible solutions to emerge (Steen, 2013). Step 2 consists of Phase 1 and Phase 2: both phases consist of co-designing with the participants with Phase 1 using media examples and Phase 2 using Audio Zoom call examples. After data collection through open ended and semi-structured interviews, thematic data analysis was conducted in Step 3 after which graphs were developed for data visualisation. Step 4 includes preparing recommendations for future work.

The majority of the interviews were conducted online for data gathering with the exception of one interview which was held in a hybrid format, with the participants in an in-person setting and the researcher facilitating online. Microsoft Teams was used as the ICT platform for online interview meetings. A consent form was read out clearly by the interviewer to the participants. They were offered sufficient time to ask questions about the content of the consent for their understanding. They could then decide whether to give their consent by speaking clearly to the recorder. All the participants gave their consent for recording their interviews. During the interview, the researchers/interviewers orally explained all the open-ended questions to the participants. The interview time for each session lasted around 2 to 3 hours with the first interview session taking 3 hours and the other 4 interview sessions taking 2 hours. The participants were English speakers, hence English was used as a primary and sole language for the interviews. To capture the data, the Teams meeting recording was used to record the hybrid and online meetings and the transcript was generated through Otter AI for data analysis. In order to secure sensitive personal information, participants' last names were not used in the transcripts of the recording.

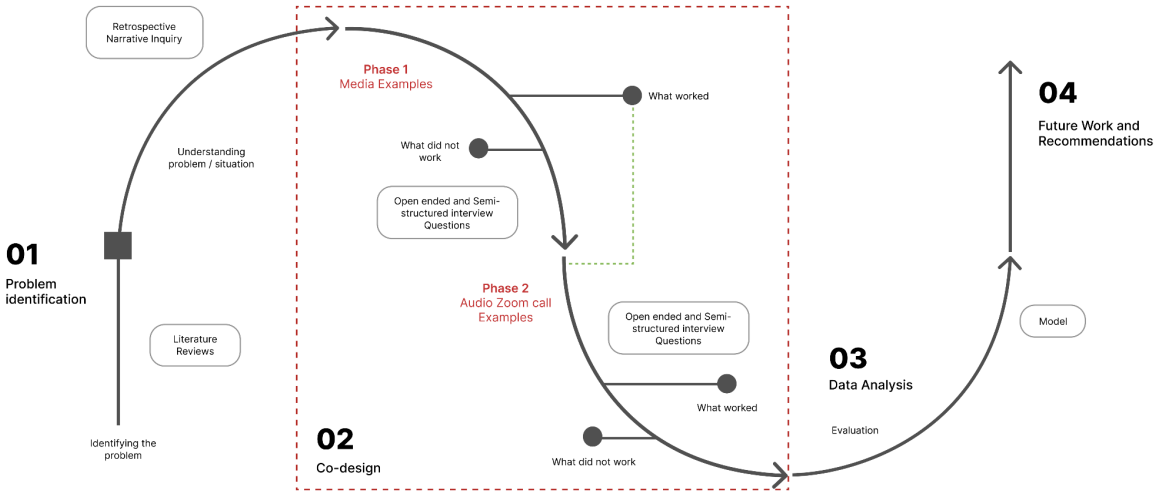


Figure 1: Study Design Methodology

## 3.2 Step 1: Open-ended and Semi-structured interviews

The first step began with semi-structured interviews. The objectives of this activity were to develop an understanding of the problem and identify pain points from the participants' experiences. They were asked of their experiences in accessing nonverbal cues when communicating with others. Since this session was part of a group interview, it was conducted in a hybrid (combined proximate and online) setting and took 3 hours.

The main themes that emerged from this was how audio cues such as tone of voice, breathing rhythm and vocal direction can signify both situational and emotional information for the participants, when communicating with close relations (e.g. family and friends). However, accessing the nonverbal cues of other people in ICT platforms was more problematic, particularly for one participant when conversing with unfamiliar people on dating websites. But to gather more data from more participants for nonverbal cues accessibility in online settings, the researcher asked semi-structured questions as an attempt to understand the problem.

However, challenges with recalling and a lack of feedback posed a problem, which is why an indirect approach, media examples, were used to not only prompt the participants in recalling experiences but also for in-situ  observations of their perception.

## 3.3 Co-design sessions

The co-design sessions were divided into two phases - Phase 1 (media examples) and Phase 2 (audio Zoom call examples). In both phases, open-ended and semi-structured interviews were conducted to delve deeper in understanding the participants' interpretations of perceived cues and their feedback on what was helping close the gaps in interpretations.

Both phases were conducted in one session, simultaneously. The co-design sessions took place online and lasted for 2 hours.

The interview protocol included two parts:

• Background: The participants were asked about what type of media they consume and why, to explore which cues are missed and what helps in interpreting the information exchanged in conversations between characters effectively.

• Nonverbal cues accessibility in online communication: The meaning of nonverbal cues was explained to the participants and then they were asked about specifications such as differences in nonverbal cue perception in online and proximate settings.

• Media and Audio Zoom Call Examples: The participants were asked what cues were perceived independently and were asked again to determine if feedback gaps occurred when silences emerged during characters' conversations, and to determine any impact on their interpretation. Additionally, participants were also asked if any other cues were assisting with closing the feedback gap.

### 3.3.1 Phase 1 - Media Examples

In Phase 1, as shown in Figure 2, familiar and unfamiliar media were used for analysing what type of audio cues already worked for the participants and which new ones would work in the unfamiliar media clips. In total, 11 media examples were used ranging from cartoons to real world media clips.

During the beginning of Phase 1, the participants were asked about their preferred choice of media and to construe the reasons behind it: how was the media edited and what type of cues were used that were perceptually and conceptually specific (Coppin, 2014) for their interpretation; what type of audio cues afforded clear interpretation, etc... After this, unfamiliar media clips were shared and as a starting point, the participants were asked to interpret interactions of the characters. From the media clips, there were two types: realistic media, for instance, slice of life genre pieces (e.g. live action movies with realistic scenes) and unrealistic media (e.g. exaggeratedly animated cartoon shows).

Realistic media was closer to real life, for making observations of realistic interactions for determining which important visual cues are missed and the impact it has on the participants' interpretation. Unrealistic media was used for not only the same reason, but also to determine the impact of other integrated alternative audio cues and observe any changes that might arise in interpretation as a result.

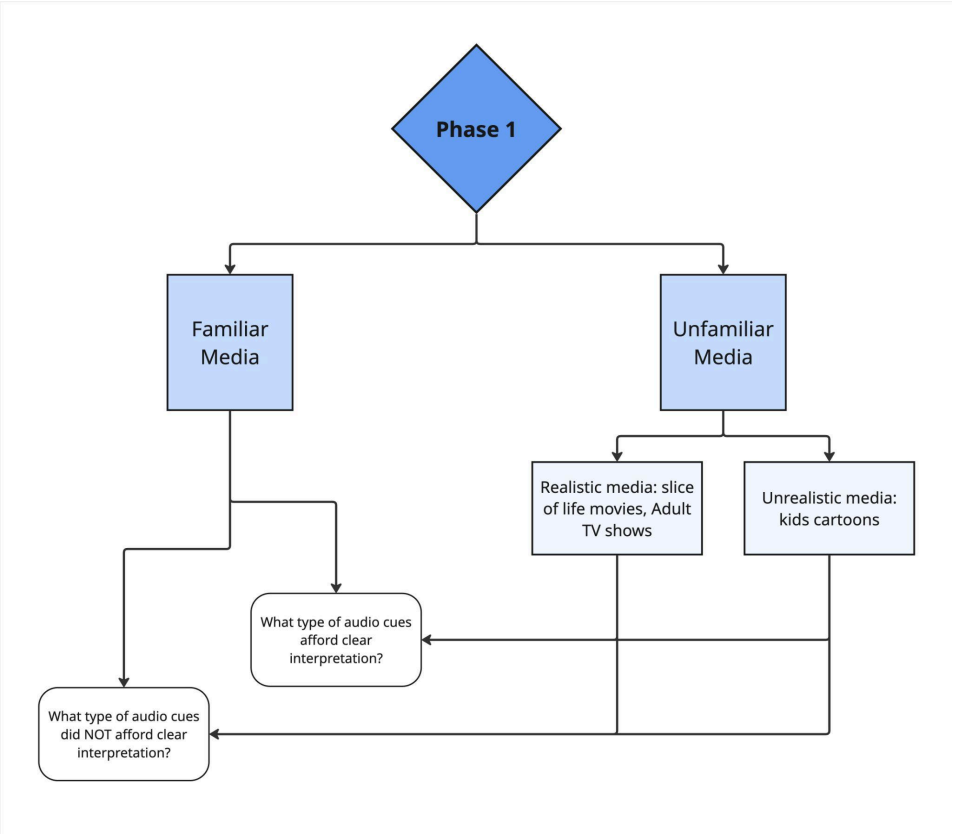


Figure 2: Phase 1 of Co-design

### 3.3.2 Phase 2 - Audio Zoom Call Examples

The purpose of audio Zoom calls was to move a step closer to realistic conversations that could happen in online meetings. Another reason for these examples was to investigate if the findings from Phase 1 supported the theory of audio cues as compensatory cues for accessing visual cues through sound.

The number of Audio Zoom calls used for the study was 4 and were not extracted from existing media. Different scripts were created on general topics of conversations for each example and an AI voice generator - Eleven Labs (*Free Text to Speech & AI Voice Generator*, 2025), was used to record the dialogs, in different voices, due to lack of time and resources for the research.

Primary audio cues studied for this phase were sound effects, descriptions and findings from this session confirmed the affordances of these audio cues as helpful for interpretation, however it also confirmed their limitations.

Table 1 enlists the media examples and audio examples that were used, along with their content details. For viewing the videos and listening to the audio examples, please see Appendix A.

| Media Examples | Context |
|---|---|
| BoJack Horseman – Herb<br><br>Category: Realistic media | BoJack comes to meet his old friend Herb in his room where Herb is on the bed, sick, and a male nurse wrapping his work before leaving. After greeting, BoJack's cheerful demeanour changes to regret and tries apologising to Herb for a past mistake. Herb calmly refuses to accept the apology before becoming angry after BoJack persists. As Herb is expressing his anger and hurt, BoJack listens with regret. |
| Avatar: The last Airbender – Sokka<br><br>Category: Unrealistic media | A dark room is shown with an eerie music and creaking sound. Sokka is facing trouble sleeping because of fear and anxiously springs in a sitting position in a defence attack stance with his sword when he hears the creak right next to him. |
| SpongeBob SquarePants -Squidward<br><br>Category: Unrealistic media | Squidward and Mr.Krabs are trying to be good servers to the restaurant's diners however Squidward is visibly more awkward and stiff because of the tension of the inspector accessing their services unbeknownst to Mr.Krabs. Squidward tries to warn Mr.Krabs of the inspector. |
| SpongeBob SquarePants -Monster<br><br>Category: Unrealistic media | SpongeBob and Patrick are standing in the middle of an area in the town for a serious matter. SpongeBob reminds Patrick to handle the serious matter, with care and sensitivity. They both yell RUN and MONSTER to everyone, resulting in the residents fleeing from the area. After everyone is gone, SpongeBob and Patrick smile at each other with a thumbs up and an uplifting sound effect. |
| Ocean's 11<br><br>Category: Realistic media | Tess is grabbing a glass while waiting on a table at a fancy restaurant for her guest. She turns in surprise to find that the man who approached her is not the person she was expecting to meet but instead her ex-lover, Danny. She is not happy to meet him and the two talk about their past; Tess visibly unhappy and Danny trying to keep cool. At the last scene, Tess shared why she was unhappy with Danny with him looking down in guilt. |
| Steven Universe – Work<br>Category: Unrealistic media | Steven and Cash are outside of an arcade. Steven is enjoying his ride on a jellyfish but Cash does not seem to be having fun. Cash begins conversing about the struggles of work, earning money and meeting people's expectations which Steven is naively unaware of. |
| Regular Show – Benson<br>Category: Realistic media | Mordecai and Rigby are dragged outside of an arcade by Benson who is visibly red with anger. Benson starts shouting angrily at them for their irresponsibility of actions that can adversely affect others like himself and result in him losing his job. |
| Kung Fu Panda – Shifu<br>Category: Unrealistic media | Shifu tells Po of his destiny of beating the villain but Po runs away because of self-doubt and inability to fulfill the mission. Shifu and Po start arguing. |

| | |
|---|---|
| Kung Fu Panda – Oogway<br>Category: Unrealistic media | Po is on top of a hill and is approached by Oogway who inquires about his sadness and anxiety. Po confesses of not meeting anyone's expectations and not being skillful enough to fulfill the mission. Oogway tries to reassure him to not be anxious about the future. |
| Teen Titans – Birthday<br>Category: Unrealistic media | Raven enters a dark room suspiciously, sensing danger but reacts shockingly from the sudden birthday surprise decorations and disappears by teleportation. The team is surprised by her reaction. Raven reappears and inquires how they all knew of her birthday date and BeastBoy confesses he secretly searched for it in her computer. |
| Teen Titans – Pancake<br>Category: Unrealistic media | Raven is flipping tar looking pancakes as Robin looks at it worriedly. The team is surprised by Raven making breakfast for them. The team reacts horribly due to the pancake's bad taste after eating except for one member who is enjoying it. |

| **Audio Examples** | **Context** |
|---|---|
| Birthday Surprise Audio | Three friends are planning a surprise birthday party for a mutual friend in an online call. One friend accidentally reveals part of the surprise too early, leading to subtle tension, then laughter, then reassurance. |
| The Quarrel Audio | Two friends quarrel with each other. One friend is angry at the other friend for getting in a fight and receiving detention. The other friend defensively confesses the reason - to stop others spreading rumours about the friend. |
| Catching Up Audio | Three friends catch up on a Friday evening Zoom call. They are talking about how their week went. Topic changes to planning a weekend retreat. Concerns arise about budget and location, but they resolve it. |
| Project Problem Audio | Three teammates on a casual call. Two members are stressed about a course but one is more sad and stressed. The other friend tries to comfort the sad friend. |

Table 2: Media and Audio Examples

## 3.4 Data Analysis

The data was analyzed following Braun and Clarke's (2006) reflexive thematic analysis approach, which involved: (a) familiarizing with the data, (b) generating coding categories and subcategories, (c) organizing data into themes, (d) reviewing and refining themes, (e) defining and naming them, and (f) selecting illustrative examples. During the initial transcript review, the concept of facework (Cupach & Metts, 1994) emerged as a guiding theoretical framework. Subsequently, key features within the data were identified to develop meaningful coding categories. These categories were then clustered into broader themes by identifying convergences in the data that revealed recurring patterns of meaning.

The main themes that surfaced from this session were interpretative gaps that emerged due to lack of access to visual cues and how participants were left with multiple assumptions in trying to fill the gaps. It also confirmed how tone of voice was effective for conveying emotional information but also music and sound effects playing a similar role as compensatory audio cues. Descriptions proved to be an effective integration for supplying and clarifying meanings. The session also highlighted limitations which will be discussed further.

# 4.0 Findings

This section presents an interpretation of aggregated findings from the open-ended and semi-structured interviews and co-design sessions using a node-link diagram for understanding nonverbal cues interpretation from media and audio Zoom call examples.

## 4.1 Phase 1 - Media Examples

The findings reveal that blind and low vision BLV individuals rely heavily on auditory cues such as tone of voice, vocal rhythm, breathing patterns, and environmental sounds to interpret nonverbal information in online settings. While tone of voice often serves as a strong alternative to visual cues, it can sometimes be ambiguous or insufficient without additional support. Media examples demonstrated how sound effects and music can either enhance or hinder cue interpretation depending on their clarity, context, and congruence with the scene. Descriptions of nonverbal cues—especially when they included the affordances or implications of gestures and expressions—were found to be particularly helpful for BLV users. Scripted Zoom call experiments further illustrated how integrating selected sound cues and descriptive narration can make real-time online communication more accessible and emotionally resonant.

A representation of the study’s findings follows, presented using node link diagrams: these node link diagrams are a qualitative analysis graph that utilise the ordinal scale of measurement for determining the interpretation (comprehensibility and incomprehensibility) of the linguistic and nonlinguistic audio cues that affected the participants' interpretation of the interactions from the media examples.

On the X-axis of the graph are displayed scenes of a media clip and on the Y-axis is a 0-5 scale from comprehensible to incomprehensible, with 0 point representing most comprehensible and 5 point representing incomprehensible. The audio cues are demonstrated using colour-coded points and lines; sound effects (blue), descriptions (pink); descriptions when video is paused (orange), music (purple) and tone of voice (green). Below the X-axis graph are highlighted details related to the audio cues; tone of voice (green) are the characters' dialogue, and type of music played in each scene is purple. In the case of descriptions, there were two different cases when using them, one was during the video

(pink) and the other was when the video was paused (orange). An explanation of the details is discussed later in the report.

All the media and recordings were played first to observe what was perceived and interpreted by the participants independent of assistive interventions such as audio descriptions. It was repeated when specific questions needed to be asked to determine their perception when a cue was expressed visually and participants did not comment about it in their summary. The clip was repeated a third time upon request by participants who were interested in further details of the events of a video and also for determining the impact of other integrated audio cues.

The purpose of the node link graph is to represent the incomprehensibility of audio cues and facilitate the recognition of patterns in the data specific to each scene, and with respect to what was working independently (e.g. sound effects) or dependently with other audio cues (e.g. sound effects with tone of voice and visual cues). Figure 3-7 visualises the findings from Phase 1, media examples while Figure 8-9 shows findings from Phase 2 - the Audio Zoom call examples.

### 4.1.1 Teen Titans Birthday video

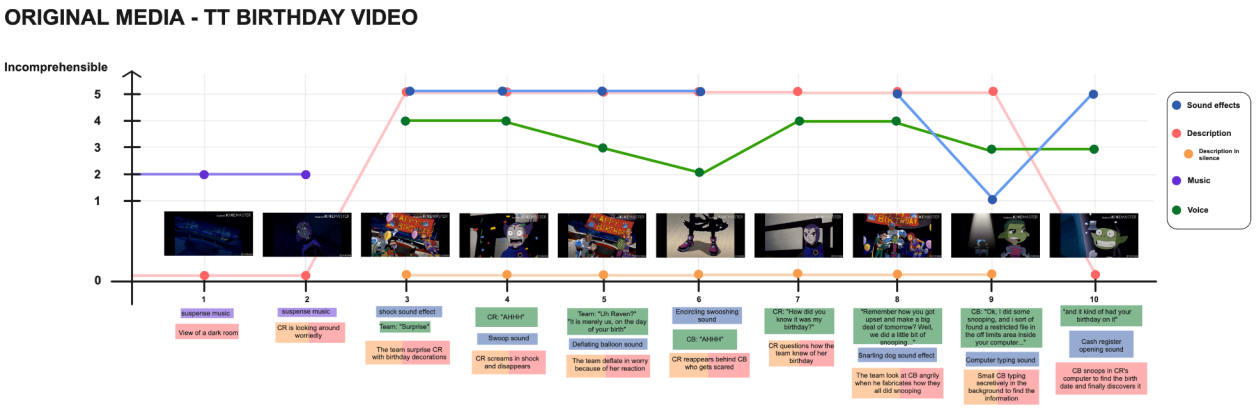


Figure 3: Teen Titans Birthday video

#### 4.1.1.1 Audio Cues: Tone of voice, Music, Sound effects, Descriptions

Figure 3, *Teen Titans* Birthday video, shows the affordances of tone of voice, music, sound effects and descriptions on the scale of incomprehensibility. The video is in the unrealistic media clips category, because they rely heavily on explicitly visual cartoon (exaggerated) language. The clip only functions when the visual language can be perceived and that it is heavily metaphorically evolved beyond the affordances of real life sensorimotor experiences.

**Music:**

The first audio that the scene started with was the music - suspenseful music which immediately alerted the participants that something suspicious was going on which was determined to be on scale 2 for incomprehensibility. Even though the participants could not determine what was actually happening during the scene, they correctly interpreted that something was suspicious from music alone. This shows that music was conceptually specific (Coppin, 2014) in providing an affective affordance; it was affording listeners an emotional or atmospheric orientation toward the unfolding scene. However, due to its abstract nature, it was not providing any background information; for example suspenseful music may afford a sense of unease or alertness, but it does not disambiguate why the scene is tense, who is involved, or what is actually happening. This causes the listener to feel something is significant but lacks the necessary information to determine exactly what that is.

**Tone of voice:**

The scale for tone of voice ranged from 2 to 4, but with most of the points close to 3 and 4 for incomprehensibility. This indicates that the participants were not able to discern much information from the tone of voice, mainly resulting from the majority of the information being expressed visually. This is another example where the cues were providing minimal information which is not conceptually specific (Coppin, 2014) enough to form a clear and concrete interpretation of what is happening and expressed. For example, in scene 7, when CR - Raven questions how the team knew the birthday's date, the participants could not determine her feelings about the birthday surprise; they were left with multiple interpretations - *was she sad? Was she upset?*

**Sound effects:**

This video was one of the two media examples that included the most sound effects. The sound effects were used for almost every type of nonverbal cue; gestures such as arms going down in scene 5, facial expressions and body language of characters looking down angrily in scene 9 as well as other types of actions such as typing on a computer keyboard in scene 10. However, as seen from the graph, the incomprehensibility scale for most of the sound effects was high, meaning participants could not determine what most of the sound effects meant. For example, in scene 5 when the characters' arms went down, even though the sound effect was similar to a balloon deflating, the participants did not know what it was signifying, showing the abstract use of sound effects - attaching sounds to reactions that do not match familiar phenomena from the real world, or pre-existing non-visual languages. Due to this abstract unrealistic use, it can only be interpreted through a perceptually specific multi-modal approach (Coppin, 2014), such as visual cues (e.g., arms going down accompanied with a surprised reaction in scene 5) to be interpreted, which is perceivable for a sighted person, but not for a blind person.

The only sound effect that was perceived immediately and understood was the computer keyboard typing sound effect in scene 9, which is why the incomprehensibility scale significantly dropped to 0 point. This is because the participants could tell since CB - Beastboy was narrating at the same time that something else was happening because of the

sound effect. These sounds support indexical semiosis, where the sound points directly to a physical action or object and are thus readily identifiable and interpretable without the need for visual support.

**Descriptions:**

Since most of the audio cues, sound effects and tone of voice were perceptually ambiguous for the participants, they wanted to know more details about the video's content. The details included information about the background setting, characters and the meaning of some of the abstract sound effects. Therefore, descriptions were provided after their inquiry. However, most of the descriptions were provided when video was paused due to the saturated presence of other audio cues which was perceptually disruptive for the participants. If the actions were described while the video was played, it would have resulted in cognitive overload (Coppin, Hung, Ingino, Quevedo, Sukhai and Syed, 2024) as seen in the graph. One of the reasons for this is determined from the line graph of both tone of voice and sound effects - their closeness shows that participants were experiencing incomprehensibility due to disrupting attention; they were trying to concentrate on the dialog, however the sound effects were distracting them. Therefore, descriptions were shared by the researchers when the video was paused.

The length and details of the descriptions was kept concise and just provided literal translations of the visual cues such as actions and what the abstract sound effects were referring to. In scene 5, literal descriptions were provided (sound effect of a deflated balloon) in addition to its affordance (a worried reaction). One participant liked the description because it provided her a clear and conceptually specific (Coppin, 2014) idea of the gesture, facial expression and their affordance. Therefore, the descriptions not only helped in clarifying the ambiguity but also in confirming the perceived audio cues - computer keyboard typing in scene 9, when CB is secretly typing in the background in order to present metaphorically as a spy finding secretive information.

#### 4.1.1.2 Conclusion and Recommendations

Music and concrete sound effects served as an effective nonlinguistic audio cue and were perceptually immediate and conceptually specific (Coppin, 2014) for informing the mood of the scene in the beginning and actions. Descriptions were perceptually and conceptually specific in informing and confirming the actions, reactions and affordances of the characters' and scenes' animation and sound. The tone of voice was ambiguous in emotional interpretation. Abstract sound effects were perceptually ambiguous and could not be interpreted.

Recommendations from participants include using a more familiar and concrete sound effect such as a "ding" or "bell" instead of a "cash register drawer opening" sound in scene 11 for conveying achievement. A potential practical implication for online meetings can be using concrete recognizable sound effects when someone is gesturing a physical action, such as silently clapping as a clapping sound, thank you bows as a cheering sound.

### 4.1.2 Ocean’s 11 video

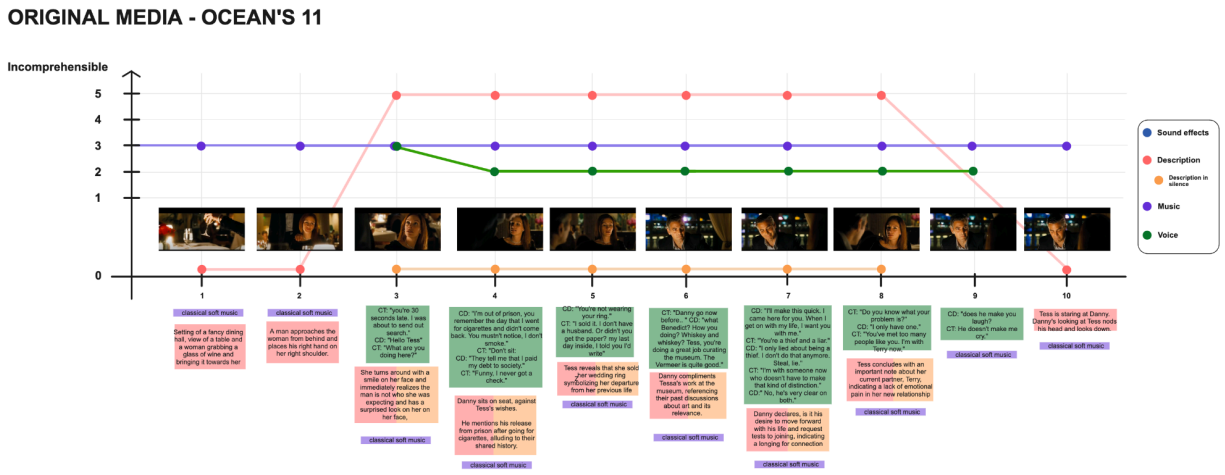


Figure 4: Ocean's 11 video

#### 4.1.2.1 Audio Cues: Tone of voice, Music, Descriptions

Figure 4 *Ocean's 11* video, shows the affordance of tone of voice, music and descriptions on the scale of incomprehensibility. The video was among the most realistic media clips, meaning the interaction was nearly the same as real life.

**Music:**

As shown in the graph, the role of music did not have a significant impact on the participants' comprehensibility. It remained constant on a scale of 3, throughout the media clip because the same music was playing (classical music) and was kept at the background of the scene, meaning that it was not giving any new information to the participants and neither was it enhancing anything conceptually.

**Tone of voice:**

Even though the characters were talking in a low tone, to some extent, the participants could understand the emotions behind the tone. For instance, from scenes 4 to 9, the participants could tell that CT - Tess was not happy talking with the other character, CD - Danny, which is why, the incomprehensibility scale decreased from 3 point to 2 point.

However, in scene 3, participants could not determine how much CT - Tess was surprised to see the other character. They could not determine if she was unpleasantly surprised, if she was in shock, or if the person was the guest she was expecting to meet. This is another

example when the tone of voice is signifying something but because it is not perceptually specific (Coppin, 2014), it created interpretive gaps and therefore, caused the participants to make assumptions to fill the gap. Another example of this is seen in scene 9, when the tone of voice became absent. The line of tone of voice in the graph shows no interpretation due to no perceptual feedback. This was a result of the characters becoming silent and only visually expressing their feelings after the conversation ended.

**Description:**

The descriptions were provided when the video was paused to avoid clashing with the dialog. Descriptions were needed for background settings and the characters, however one participant wanted to know what was being expressed by CD - Danny in scene 10. She was uncertain due to the multiple assumptions that were made in an attempt to know for certain what happened, therefore a description describing CD's literal visual reaction was described. The description further added the facial expression's affordance to which the participant immediately agreed with, meaning that she was conceptually aware of what the visual cue meant. This was further confirmed when the same description was provided to another participant for whom it was confirming her intended interpretation which was CD expressing guilt. This shows when silences emerge after a conversation, some BLV individuals have an idea as to what could be happening but for some, it becomes a disruption of feedback, meaning they are not receiving any information. Audio cues such as descriptions either confirm their idea or inform them.

**Sound effects:**

While there were no sound effects used in the video, one of the participants used this clip as an example where sound effects can be ineffective and unnecessary. She shared that because sound effects are generally used in cartoons or for humour purposes, it would be inappropriate or cartoony in serious conversations such as in this clip. Thus sound effects would feel misplaced and disrupt the flow of the conversation as well as the meaning making process.

#### 4.1.2.2 Conclusion and Recommendations

In conclusion, tone of voice was not perceptually specific (Coppin, 2014) but when tone of voice became absent, literal concise descriptions afforded immediate perception and comprehension. The music vaguely conveyed information about the setting, at the beginning of the scene.

One participant did not recommend using sound effects such as a pop or bell in these types of serious conversation as it would be unnecessary and inappropriate. The media example also revealed the importance of gaps serving as a space for integration descriptions and thus resulting in effective interpretation. A practical implication that emerges from this is how concise and direct descriptions can be provided when two people stop talking and their facial expressions are described to the BLV individual or when the online space falls in silence

when people stop talking momentarily and their facial expressions or gestures can be described if they are visually expressing something.

### 4.1.3 SpongeBob Monster video

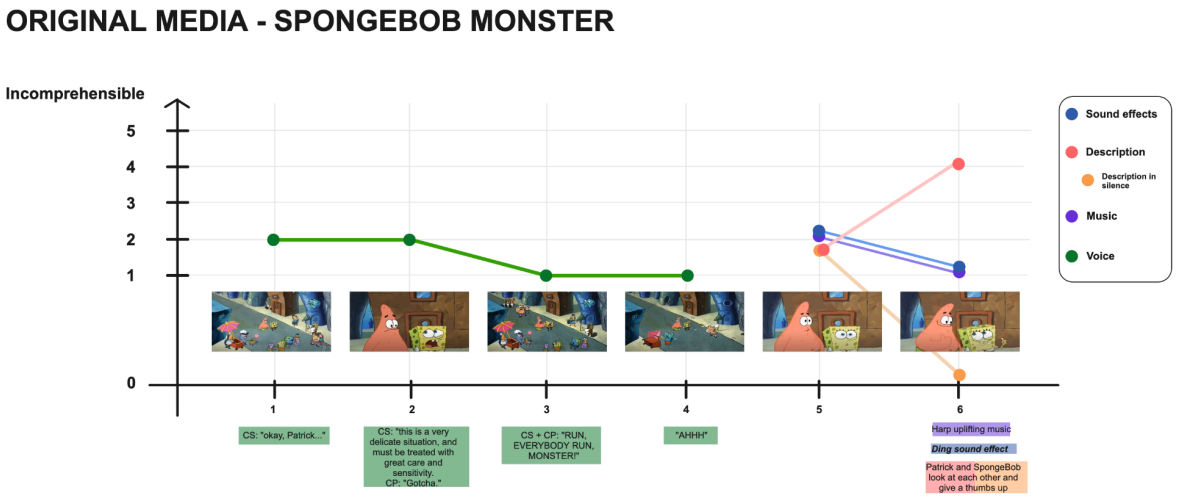


Figure 5: SpongeBob Monster video

#### 4.1.3.1 Audio Cues: Tone of voice, Music with Sound effects

Figure 5 SpongeBob Monster video, shows the affordance of tone of voice, music with sound effects and descriptions on the scale of incomprehensibility. The video was in unrealistic media clips.

**Tone of voice:**

The tone of voice was limited in this clip however, due to the situation, it was giving some idea to the participants as to what is happening. However, similar to the *Ocean's 11* video in scene 9, the characters stopped talking and so the tone of voice became absent. The difference comes from the integration of sound effect and music exhibited in sync with the characters' visual cues; facial expression (smile) and gesture (thumbs up).

**Music:**

The music in this video was played simultaneously with the sound effect; a harp and uplifting chime which was then followed by a 'ding' sound effect. The music served to inform the participants that something positive happened for the characters and its continuation upon the sound effect's addition further enhanced the positive emotion. This shows an example of contextual layering where music acted as a contextual scaffolding audio cue and the 'ding' afterwards enriched it as an achievement.

However, there were two limitations: the music's abstractness, and its context-dependency. In the case of abstractness, one participant could not determine what the music was signifying and considered it as misplaced and unnecessary - it reminded her of a commercial ad and thus did not inform her of any information it was signifying. The other, like the sound effects, is its dependence upon other cues. Independently, due to its abstractness and flexibility, this music can signify any type of action, which is why it needs to be anchored with a context (content of the video) and the visual cues (smile and thumbs up) to signify a conceptually specific (Coppin, 2014) meaning (achievement).

**Sound effects:**

The only sound effect used was at the end in scene 6, when the characters are expressing their joy visually. The chime followed by a 'ding' sound effect was effectively affording emotional information; a happy feeling and a sense of achievement, as mentioned by participants (e.g. The dialog "we did it"). However, an interesting finding from this is that while the emotion was interpreted, the visual cues were not perceived and thus were not identifiable when the participants were asked what was depicted. This shows an example of affective affordance, where audio cues like sound effect was serving as a nonlinguistic compensatory cue which effectively afforded what the gesture and facial expression meant even though they could not be identified. In this way, it was reducing incomprehensibility from 2 point to 1, by providing information

However, because these sound effects are not actual sounds of actual visual cues expressed in real life, they are abstract and not universally interpreted and are conceptually ambiguous for some people. This was confirmed when one participant did not know what the nonlinguistic audio cues meant and not only was unable to identify the visual cues but also could not determine their meanings. Another significant limitation of sound effects such as a 'ding' is that it is significantly more context dependent than music. A 'ding' can be applied to any reaction and action and while music can have some level of being conceptually specific (Coppin, 2014) (uplifting chime meaning positive feeling), a 'ding' is more abstract and thus arbitrary in interpretation.

**Description:**

Description was primarily required for the background setting, because the participants could not determine the details about the situation. The spatial audio was giving them a limited idea as to where the characters could be but it was not confirmable. Only one participant wanted a description of what happened at the end because she could not determine what the sound effects and music meant. The description was immediately interpretable for her and served as a conceptually specific (Coppin, 2014) audio cue, filling the feedback gap that emerged.

#### 4.1.3.2 Conclusion and Recommendations

In conclusion, for some people, music can be conceptually specific (Coppin, 2014) for affording emotions, depending on context. Abstract sound effects do not afford any

interpretation unless it is anchored with other non-linguistic compensatory cues such as music. However, for some people, independent or combined, abstract non-linguistic cues are not perceptually and conceptually specific (Coppin, 2014). Descriptions with affordances are immediately perceived.

Recommendations are based on the effectiveness from the combined use of music and sound effects in conveying emotional information. Music and abstract sound effects (depending on genre of music) can be used together to convey different meanings of gestures and facial expressions. For example, for a sad frown face, music like a low, melancholic piano combined with a glum sound can potentially convey sorrow and a playful, mischievous smile can be conveyed using something similar to a slide whistle up sound effect paired with a bouncy or a quirky tune to signify silliness or lightheartedness

### 4.1.4 BoJack Horseman video

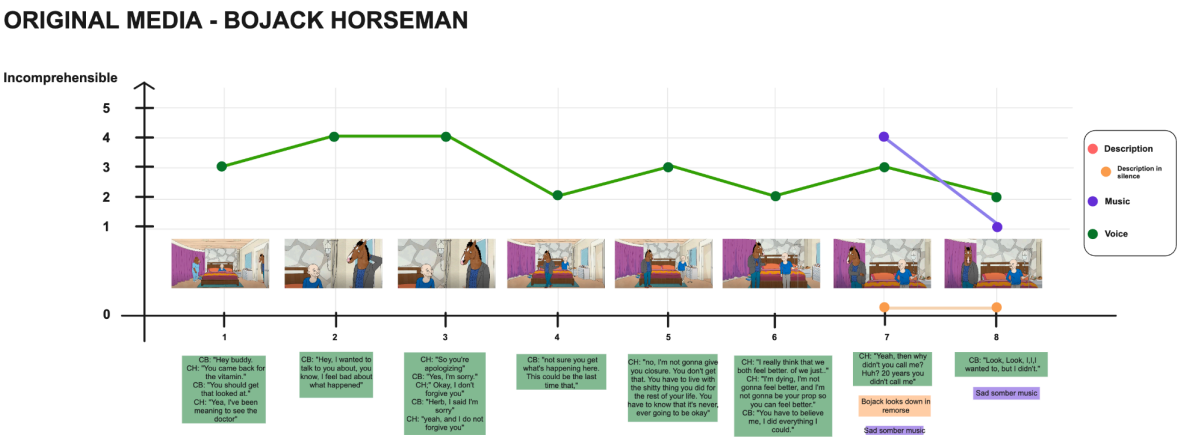


Figure 6: BoJack Horseman video

#### 4.1.4.1 Audio Cues: Tone of voice, Music, Descriptions

Figure 6 *BoJack Horseman* video, shows the affordance of tone of voice, music and descriptions on the scale of incomprehensibility. The video is in the realistic media clips category, meaning the animated characters' nonverbal cues were similar to real life .

**Tone of voice:**

As shown from the graph, throughout the video, the tone of voice remains between 2 and 4 on the scale of incomprehensibility with the majority of the points occupying the 3 and 4 margin. The reason for this is because the participants could not tell the characters' emotions from their tone of voice. For example, it was not perceptually clear enough for them to

determine how upset CH - Herb was or how guilty CB - BoJack felt. This is also an example of how tone of voice was resulting in the intended meaning remaining partially ambiguous or open to multiple interpretations, (how upset or how guilty they are). In such cases, supplementary cues such as music acted as a disambiguating uncertain perceptions, therefore enhancing the interpretation of characters' feelings for the participants.

**Music:**

As shown in the graph, the music decreased the incomprehensibility when interpreting significantly, from 4 point to 3 point on the scale from scene 7 to scene 8, and therefore helped clarify the intended meaning and resolve interpretive uncertainty. The integration of music, as mentioned by participants, helped form an understanding of the whole scene, when one of the characters was expressing his hurt verbally and the other was expressing his guilt and sadness visually, the music evoked a feeling of sadness from both the characters. The participants could not see the visual cues of the other character, but they received minimal feedback of what he could be feeling through the music. This also shows the independence of the music's affordance – where descriptions are not needed for clear interpretation as music is sufficient. This also shows that music provides an additional, nonverbal layer of meaning that interacts with verbal or visual information. Even when emotional cues are present visually or verbally, music amplifies, deepens, or nuances those cues, creating multimodal emotional convergence.

**Description:**

Descriptions acted as another compensatory cue which aided the participants to confirm their interpretation after listening to the dialog and music. The descriptions also acted as confirmatory semiotic scaffolds, offering explicit feedback loops that informed the character's gesture and confirmed interpreted emotion.

#### 4.1.4.2 Conclusion and Recommendations

In conclusion, music was conceptually specific (Coppin, 2014) in determining the emotions and overall mood, but not for indexical specification of the gesture and the descriptions provided were conceptually specific for concrete affordance framing in understanding its meaning.

### 4.1.5 Kung Fu Panda Shifu video

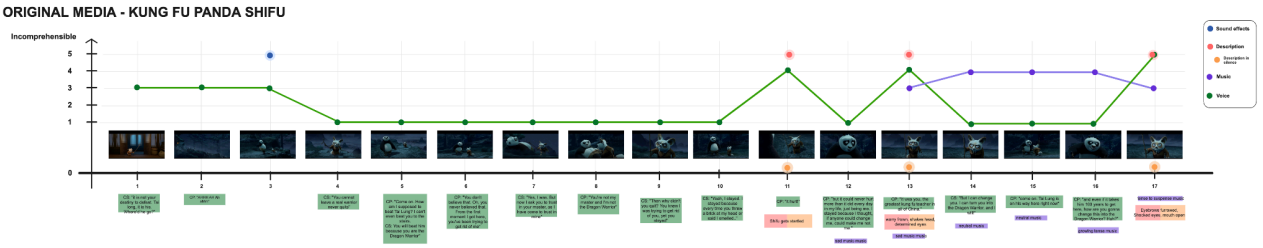


Figure 7: Kung Fu Panda Shifu video

#### 4.1.5.1 Audio Cues: Tone of voice, Music, Descriptions

Figure 7 *Kung Fu Panda Shifu* video, shows the affordance of tone of voice, music and descriptions on the scale of incomprehensibility. The video was in unrealistic media clips.

**Tone of voice:**

The participants noted the tone of voice being very expressive; it was perceptually clear for them in understanding the characters' feelings; anger, frustration, hurt, assertiveness, sadness, etc... It is for this reason that the tone of voice remained at a constant level of comprehensibility throughout the video, from scenes 4 to 10, at the 1 level. The only time it changed, it moved to the 3 position, which was when the other character became silent. When asked, the participants did not feel the need for the visual cues to be described, particularly due to the music played in the background which was setting the mood of both the characters and the overall conversation.

**Music:**

As shared by a participant, in the *Kung Fu Panda Shifu* scene, in many different moments, CP - Po character was expressing his hurt and sadness while CS - Shifu was listening and expressing his thoughts and feelings visually which created a silence and where the participants were misinterpreting what CS - Shifu was expressing. While the participant could not clearly interpret CS's expressions, she could understand what CS was feeling because of the music as shown in graph, with an decrease in incomprehensibility from 4 to 3 point – this is when the flow of music was changing from sad to hurt and then suspense, which shows the flexibility of music and how its non-linguistic nature can be effectively used for affording emotions without clashing with linguistic sounds - dialogs. It was enhancing the hurt in CP's voice which then smoothly transitioned to suspense to signify the shock and speechless reaction of CS. The music played in the silence effectively afforded what CS was thinking: as shared by the participant, "not sure how he's gonna do it".

**Description:**

In this video, participants were provided with literal descriptions of visual cues. This was to determine if alternative and perceptually specific (Coppin, 2014) audio cues were conceptually specific (Coppin, 2014) for them for interpreting the meaning. As shown in the graph, the concise linguistic descriptions were comprehensible for the participants. It was acting as a surplus, meaning that even though the participants did not need descriptions, it was adding an extra layer for interpretation - confirming what they perceived.

However, despite the descriptions being concise and literal, the video still needed to be paused to allow for clear interpretation because the graph demonstrates the same concise descriptions reaching to point 5 in incomprehensibility, because it was clashing with the dialog and thus, could not be perceived at all.

#### 4.1.5.2 Conclusion

In conclusion, the music was acting as an alternative compensatory nonlinguistic cue in providing conceptually specific (Coppin, 2014) information about the emotions and thoughts. Concise and literal descriptions were conceptually specific and provided concrete affordance framing in understanding its meaning with tone of voice being perceptually specific (Coppin, 2014).

In practical terms, the duration of music used to convey emotional information does not need to be lengthy. Findings from this study, as well as broader media examples, suggest that music inherently carries concrete emotional meaning—such as how upbeat rhythms are intuitively associated with joy or excitement. Therefore, even brief segments of music can effectively communicate emotional tone and set the mood. Importantly, this is not limited to enhancing the experience for those present in a live meeting; it also serves as a meaningful emotional cue for BLV individuals joining later, functioning not only as a mood enhancer but as a standalone supplier of emotional context.
For instance, a short, cheerful jingle played as someone shares good news during a meeting can instantly signal a positive atmosphere. If a BLV participant is joining after the moment has passed, that snippet of music can help them grasp the joyful tone of the interaction, even without seeing facial expressions or body language.

### 4.1.6 Summary

The table lists the overall findings from Phase 1 of  affordances and limitations of perceived audio cues.

Findings reveal the interpretation gaps that emerged due to lack of perceptual access to visual cues expressed by characters. Observations of how tone of voice acted as a compensatory cue for signifying emotional information however its limitations were also identified. Literal concise descriptions acted as direct translators of visual cues and supplied information that was missed due to perceptual inaccessibility. They were understood by the

participants however, detailed descriptions which also provided the meanings of translated visual cues, were also enjoyed. Issues with descriptions emerged when it clashed with other linguistic cues like speech and abstract sound effects.

Abstract sound effects were not immediately perceived unlike concrete sound effects that were instantly understood. The participants could not identify signified meaning of the abstract sound effects which resulted in multiple interpretations. Hence, when speech and abstract sound effects coincided, it was resulting in cognitive overload (Coppin, Hung, Ingino, Quevedo, Sukhai and Syed, 2024)

Lastly, concrete sound effects were more effective for indexical representation of actions, some abstract sound effects and music were more effective emotional interpretation and were either supplying an emotional layer or enhancing it

| | | Affordances | Limitations |
|---|---|---|---|
| **Descriptions** | Literal, concise | **Conceptually specific - Symbolic**<br>Direct affordance indicators, allowing the perceiver to reconstruct the nonverbal event mentally affording clear, easy-to-access meaning.<br><br>Example: Nods his head, looks down [Scene 11 - Ocean's 11 video] | **Conceptually ambiguous - Symbolic**<br>Can result in interpretive ambiguity due to semiotic underdetermination - they provide incomplete or ambiguous meaning for someone who is not familiar with the associated meaning<br><br>Example: Shrugs, Eye roll |
| | Detailed with affordances | **Conceptually specific - Symbolic**<br>Multi-layered meaning, where the action or nonverbal (visual) cue is described in addition to its afforded meaning.<br><br>Example: Looks down in remorse - [Scene 7 - Bojack Horseman] | **Cognitive overload - Symbolic**<br>Excessive detail can result in cognitive overload; caused by redundancy therefore, unnecessary. It can be distracting since it not informing or enhancing perceived information and can clash with other linguistic audio cues (dialog) and non-linguistic audio cues (sound effects, music). |
| **Sound effects** | Concrete | **Iconicity - Perceptually immediate - Conceptually specific - Situational specificity**<br>Exhibit high iconicity - maintained a direct, perceptually motivated relationship with the event or object they represent. Direct indexical interpretation, identifying clear physical actions non-linguistically<br><br>Example: Computer typing sound - [Scene 9 - Teen Titans Birthday video] | **Context-dependent - Conceptually ambiguous**<br>Needs to be applied according to context and anchored with other cues for effective construction multi-modal layered interpretation. |
| | Abstract | **Symbolic - Affective Affordance**<br>Exhibit greater semiotic arbitrariness - the sound effect does not identify the action directly. Instead, it affords an emotional atmosphere or interpersonal cues rather than discrete actions.<br><br>Example: Ding sound - [Scene 5 - SpongeBob Monster video] | **Symbolic - Perceptually and Conceptually ambiguous - cross-modal anchoring**<br>Results in semiotic dissonance - a breakdown in meaning-making where it does not align with, enhance, or clarify the visual or narrative content, leading to cognitive overload, distraction, clashing and therefore unnecessary.<br><br>Example: Chime, ding sound |
| **Music** | | **Symbolic - Affective Affordance**<br>Operates primarily through **affective affordances** — it evokes general emotional atmospheres (sad, suspense, happy) rather than conveying specific, propositional meaning. It can also inform the mood of a scene.<br><br>Example: Sad somber music - [Scene 7 - BoJack Horseman video] | **Symbolic - Context dependent - Situational ambiguity**<br>Can not specify what is happening when enhancing the feeling or mood of the scene (semiotic underdetermination).<br><br>Example: Sad somber music (sad feeling but gesture unidentifiable) - [Scene 7 - BoJack Horseman video] |
| **Tone of Voice** | | **Perceptually specific - Affective Affordance**<br>functions as a **paralinguistic semiotic resource**, allowing for the interpretation of emotional states, intentions, and interpersonal attitudes. provides affective affordance, enabling listeners to perceive subtle emotional and psychological meanings without the need for additional descriptive input.<br><br>Example: Anger, Frustration, Hurt, Sad emotions in dialog - [Kung Fu Panda Shifu video] | **Ambiguity resulting in multiple interpretations**<br>When tone of voice is unclear, the intended meaning remains partially ambiguous or open to multiple interpretations.<br><br>Example: How sad, how guilty? - [Bojack Horseman video] |

Table 3: Phase 1 - Media Example Findings

## 4.2 Phase 2 - Audio Zoom Call Examples

The purpose of the Audio Zoom calls was to transition closer to real life online meetings and determine the effectiveness of the audio cues as a compensatory or enhancing cue. This was another way to observe the effectiveness of tone of voice and to identify the impact of findings from Phase 1. The audio cues used and studied for this phase were sound effects (abstract), descriptions (literal, concise).

### 4.2.1 Birthday Surprise Audio Zoom Call

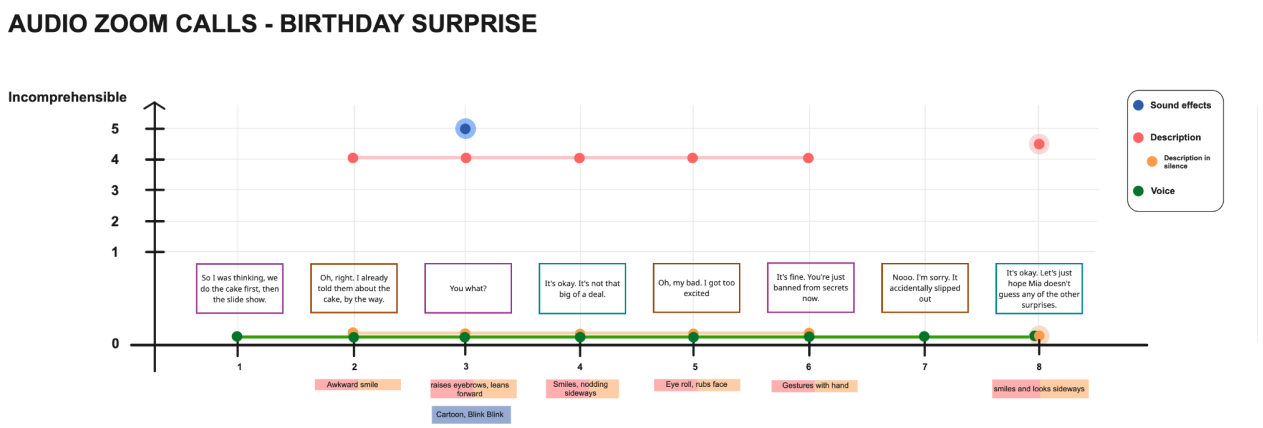


Figure 8: Audio Zoom call - Birthday Surprise

#### 4.2.1.1 Results

Figure 8 Audio Zoom call - *Birthday Surprise*, shows the affordance of tone of voice, sound effects and literal concise descriptions on the scale of incomprehensibility.

**Tone of voice:**

The tone of voice was perceptually clear and specific in interpreting information about the content and emotions. Due to the recording's short length, the conversation was easy to follow and the participants did not feel the need to have any additional audio cues. For this reason, the entire recording was comprehensible for the participants as displayed on the 0 position for incomprehensibility.

**Sound effects:**

The only sound effect used was a twang sound indicating the expression of eyes blinking. The twang is a short, sharp, resonating sound, particularly the sound of a stringed instrument being plucked. The graph shows that it was incomprehensible (point 5) because the participants reported it being uninterpretable. The reason for this was due to its abstract nature, meaning it was conceptually ambiguous and thus the participants were unable to determine what it was signifying.

**Description:**

The descriptions were not needed by the participants, however they were open to observe any impact it had on their interpretation upon perceiving it. The participants responded with positive feedback and enjoyed how descriptions were adding an extra layer of meaning to the conversations. This shows that descriptions (literal and concise describing visual cues) served as a cross sensory redundancy cue (Coppin, Hung, Ingino, Quevedo, Sukhai and Syed, 2024)

- even though it was not needed, it was enjoyed and acted as an enhancer to the meaning making process.

#### 4.2.1.2 Conclusion and Recommendation

In conclusion, the tone of voice was conceptually specific (Coppin, 2014) for content and emotional interpretation. Abstract sound effects were conceptually ambiguous and led to no interpretation. Literal concise descriptions of visual cues act as confirmatory scaffolding, further enhancing the interpretation and acting as a cross modal redundant signifier (Coppin et al., 2024; Part G.1.1).

A future implication of this can be when in an online meeting, a person rolls their eyes in response to a sarcastic comment. A literal description like "[Person rolls their eyes]" can be added through something similar to live transcription. While a BLV participant might already infer the sarcasm from the tone of voice, this concise visual translation confirms and enriches their understanding by offering a layer of social nuance reinforcing that the expression was not just verbal but also physical.

### 4.2.2 Catching Up Audio Zoom Call

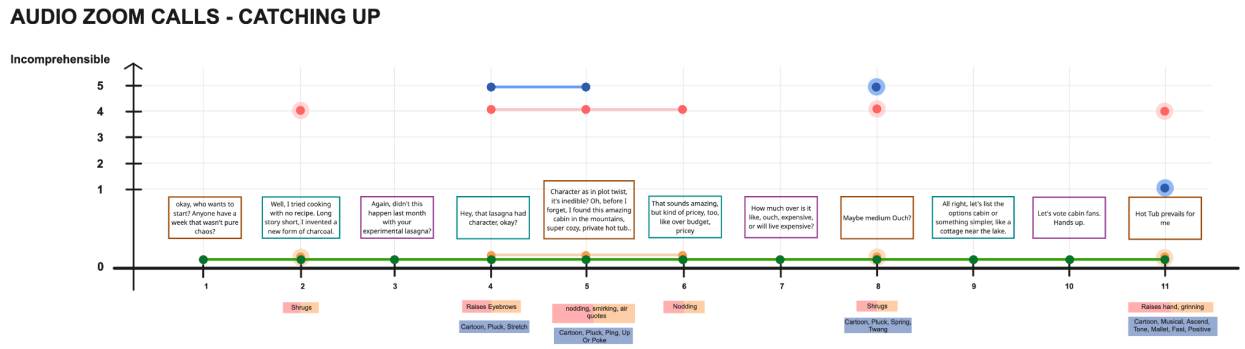


Figure 9: Audio Zoom call - Catching up

#### 4.2.2.1 Results

Figure 9 Audio Zoom call - *Catching up*, shows the affordance of tone of voice, sound effects and literal concise descriptions on the scale of incomprehensibility.

**Tone of voice:**

The tone of voice was perceptually clear and specific in interpreting information about the content and emotions, in this recording as well. The graph shows the tone of voice at the 0 position for incomprehensibility, meaning it was completely comprehensible.

**Sound effects:**

Four sound effects were used for this recording, in scenes 4,5, 8 and 11 - the twang sound effect which differed in tones and rhythms. In scene 4, a cartoon, pluck, stretch sound for eyebrows raising - a similar way how they are stretched upwards too. In scene 5, a cartoon, pluck, ping up or poke sound effect was used to demonstrate the air quotes. The sounds in this audio are repeated twice. In scene 8, a longer sound effect was used; cartoon, inward spring sound. which signified a nervous shrug - a similar way of how a person's body shrinks when nervously shrugging when unsure of something. In scene 11, a chime ascended sound was used to signify a positive gesture (a thumb up).

With the exception of the chime sound in scene 11, the rest of the sound effects were conceptually ambiguous and the participants could not determine what they were signifying. The graph shows that in scenes 4,5 and 8, the sound effects were at the 5 position due to their incomprehensibility and the participants noted how they were distracting and resulting in modal noise meaning that this was only acting as an extraneous material which was not enhancing, clarifying, or deepening the meaning and was diluting their focus on the dialog.

**Description:**

The descriptions were not needed by the participants in this audio example. This is because the tone of voice was expressive independently and therefore, the participants did not need external linguistic assistance for interpreting the information. This was further proved when descriptions were played as part of procedure and the participants confirmed that it was necessary. However there were no reports of description being redundant.

#### 4.2.2.2 Conclusion

In conclusion, the tone of voice was conceptually specific (Coppin, 2014) for content and emotional interpretation. Abstract sound effects were conceptually ambiguous and led to no interpretation, except for one abstract sound effect (chime) that had some level of concreteness in meaning - was more conceptually specific (positive feeling). Literal concise descriptions of visual cues were not needed and had no impact in interpretation.

### 4.2.3 Summary

The tone of voice was clear in both the examples and was sufficient for emotional interpretation. Literal concise descriptions and abstract sound effects were variably effective - descriptions were effective in supplying information as literal auditory translations of visual cues and abstract sound effects for enhancing the feeling already perceived from tone of voice. However, since the recordings were short and easier to perceive cues, these audio cues were not specifically required.

In *Birthday Surprise* audio, literal concise descriptions were conceptually clear and acted as a cross sensory redundant signifier (Coppin, Hung, Ingino, Quevedo, Sukhai and Syed,

2024); they were enjoyed by the participants. The abstract sound effect (e.g., Abstract Blinking sound) could not be interpreted and was distracting and unnecessary.

In Catching Up audio, literal concise descriptions were not needed and one participant found the abstract chime sound effect as a positive enhancer but the other participant found it unnecessary and unmeaningful.

Table 3 summarizes the findings from Phase 2 and shows the affordances and limitations of investigated audio cues

| | | Affordances | Limitations |
|---|---|---|---|
| **Descriptions** | Literal, concise | **Conceptually specific - Direct**<br>Participants received verbal access to expressed visual cues. It added an extra layer of meaning which was enjoyed by the participants.<br><br>Example: Awkward smile, Eye brow raised, leans forward [Scene - Audio Zoom call Birthday Surprise] | **Not immediately needed**<br>Acting as a surplus signifier means since they are not functionally required, they can become excessive in some situations and thus not needed to prevent cognitive overload<br><br>Example: Catching up |
| | Detailed with affordances | | |
| **Sound effects** | Concrete | | |
| | Abstract | **Symbolic - Affective Affordance**<br>Enhanced the positive feeling of the conversation. Added an extra layer of positive feeling after the tone of voice.<br><br>Example: Chime sound effect - [Scene 11 - Audio Zoom call, Catching up] | **Symbolic - Perceptually and Conceptually ambiguous - Unnecessary**<br>Possesses mixed reviews due to its abstract nature. Someone who is conceptually familiar with the abstract sound effect's affordance, enjoyed it. For someone who is not, it was distracting and unncessary.<br><br>Example: Chime sound |
| **Music** | | | |
| **Tone of Voice** | | **Perceptually specific - Affective Affordance**<br>Allowed for clear interpretation of emotional states, intentions, and interpersonal attitudes. The tone was clear and contextually grounded, effectively achieving **modal sufficiency,** meaning the auditory modality alone provides enough information to construct a coherent interpretation.<br><br>Example: Catching Up and Birthday Surprise Audio Zoom Calls | |

Table 4: Phase 2 - Audio Zoom Call Example Findings

# 5.0 Discussion

This section discusses the affordances and limitations of audio cues in depth and their role as an alternative mode to visual cues for supplying, clarifying or enhancing situational and emotional information and as a compensatory cue as well.

Through examining scenes from entertainment media and simulating real-time online conversations via scripted Zoom calls, this research investigated how BLV individuals perceive and interpret nonverbal cues in the absence of visual information. The study identified key patterns in how different types of auditory information such as tone of voice, sound effects, music, and descriptive narration, either support or hinder access to social and emotional meaning.

Additionally, understanding how the findings can be applied in practice is also important. This is why a specific hypothetical scenario motivated by Lee, Sukhai, Coppin (2022)'s paper will be used as an example (a colleague entering their manager's office). The example is changed from in-person to an online setting and instead of the colleague entering the office, a sighted colleague and a BLV colleague are joining late to an online video call meeting with their manager. The manager had joined for some time and the sighted colleague notices the manager's stressed mood nonverbally but for the BLV colleague, unless the manager starts speaking, they cannot detect this mood.

## 5.1 Visual Cues

BLV participants experienced modal inaccessibility, where visually expressed cues  such as facial expressions, gestures, or body movements — are inaccessible due to the reliance on the visual modality. This leads to critical signs that would typically guide understanding are not perceived, resulting in a gap in meaning-making.

As the interaction continued and silence occured (e.g., in *Ocean's 11* video, scene 10), the absence of sensory feedback triggered a perceptual gap: the individual senses that meaning is unfolding but lacks the perceptual information to interpret it. In response, the perceiver generates hypotheses to fill in missing meaning based on limited cues and prior experience.

From the media examples and visual cues expressed by characters, the participants confirmed how these gaps in perception and interpretation can arise if someone was expressing the same cues such as *looking down in regret* from the *Bojack Horseman* video clip (scene 7) and *nodding and looking down* in *Ocean's 11* video (scene 9). This can be applied in the online meeting example when the BLV colleague would miss the emotional information expressed from the manager's visual cues.

## 5.2 Tone of Voice

The tone of voice (paralinguistic cues) functions as a paralinguistic semiotic resource, allowing for the interpretation of emotional states, intentions, and interpersonal attitudes (Mamurova, 2024). Through auditory semiosis, the tone of voice provides a rich affective affordance, enabling listeners to perceive subtle emotional and psychological meanings without the need for additional descriptive input. This is confirmed by Lopez (2022) how participants experience vococentrism by having a precedence of the voice over other sounds, resulting in more focus on what was being said than what was being denoted through other sounds, which was specifically highlighted from the findings of the *Teen Titans Birthday* video clip.

When tone is clear and contextually grounded, it effectively achieves modal sufficiency, meaning the auditory modality alone provides enough information to construct a coherent

interpretation, making extra descriptive elaboration unnecessary. This was confirmed from both the *Kung Fu Panda Shifu* video clip in Phase 1 and *Catching Up* audio Zoom call examples in Phase 2. Similarly, the tone can also signify tension and fatigue of the manager in the online meeting example, if it is clear.

While tone of voice often serves as an effective alternative mode for conveying meaning, layering it with different audio cues became necessary to provide conceptually specific (Coppin, 2014) information which was either informing or confirming the uncertain interpretation; this was evident from the *Bojack Horseman* video clip (Scene 7) when music helped convey the emotions of the characters, while tone of voice could not. This also indicates that there are situations in which tone of voice was resulting in intended meaning remaining partially ambiguous or open to multiple interpretations, and therefore compensatory audio cues became necessary to ground the conceptual understanding and reduce uncertainty. For the sighted colleague, if the manager started speaking and as he is conversing, the tone of voice (tensed) and visual cues (slumped posture, rubbing forehead) would signify the stress more concretely.

## 5.3 Sound Effects

Sound effects' abilities to convey meaning is classified based on the degree of iconicity and arbitrariness inherent in the sound's relation to its source. Concrete sound effects exhibit high iconicity - they maintain a direct, perceptually motivated relationship with the event or object they represent. This was observed in the scene in which a computer keyboard typing sound was solely identified and helped in giving situational context. From the constructivist viewpoint, the magnetism of synchronous visual and auditory objects is an expression of the inherited causal thinking: whenever two events occur simultaneously and roughly close to each other in space (hence spatio-temporal congruency), perception compellingly creates a causal connection (Görne, 2019; Riedl, 2016). In this way, concrete sound effects also afford, "natural meaning" Grice (1957) - where the signifier (e.g., computer keyboard typing sound) was signifying the action of someone typing on a computer.

Abstract sound effects, on the other hand, exhibit greater semiotic arbitrariness. Though produced through real-world means (such as a chime or bell ding), their use becomes metonymic or symbolic, detached from their physical origin. In cases such as a chime accompanying a character's smile as seen from the *SpongeBob Monster* video clip, the sound effect does not identify the action directly. In another example, the *Catching Up* Audio Zoom call also resulted in the same gap in directly identifying an action. Instead, it afforded an interpretation of positive emotional meaning, signaling something uplifting, friendly, or good. This abstract sound use relies on synchronisation: it becomes meaningful when paired with visuals (e.g. a smile) (Chion, 2001). Thus, while concrete sounds support event identification, abstract sounds function primarily through affective affordance, conveying emotional atmosphere or interpersonal cues rather than discrete actions.

Ultimately, sound effects in media operate along a continuum: from direct action signification (e.g., computer keyboard typing) to emotionally enhancement (e.g., cash register ding indicating a jackpot), depending on the level of iconicity and contextual integration within multimodal meaning-making. In the online meeting example, abstract sound effects can be used to signify tension through non-linguistic sounds. For example, a repetitive ticking or pulsing tone can evoke mental pressure or time urgency based on the manager's actions, for instance, if they are being rushed through their work.

However, this also resulted in a breakdown in meaning-making where auditory cues do not align with, enhance, or clarify the visual or narrative content, leading to cognitive conflict or distraction. Additionally, these sound effects exhibit few affordances as well - their auditory qualities do not afford intuitive interpretation (e.g., snarling dog sound for angry reaction) or immediate action possibilities (e.g., swoop sound for disappearing), making them difficult for the perceiver to integrate into the unfolding scene. Where effective sound design would offer indexical or affective anchoring to the narrative (pointing to actions or emotional tones), here the sounds remain resistant to straightforward interpretation, and thus failing to contribute to multimodal meaning-making.

## 5.4 Descriptions

The descriptions acted as a sentential representation for conveying situational and emotional information in a logical sequence (Larkin and Simon 1987; Lee, Sukhai and Coppin, 2022) which was expressed visually by the characters and translated through arbitrary linguistic signs (words) to the participants. When sensory signs (e.g., sound effects, body movements) provided incomplete or ambiguous meanings, descriptions were suited to intervene by providing conceptually specific (Coppin, 2014) information regarding the setting or reactions, with and without their meanings. This anchoring clarifies who, what, where, or why something is happening, supporting the completion of the meaning-making process. This was strongly evident in the findings of the *Teen Titans* and *Ocean's 11* video clips (see Figure 3 and 4).

Descriptions also enriched interpretations: through words, they expanded the meaning potential beyond what is immediately perceptible. For example, from the *Ocean's 11* video clip, "*looks down*" might not reveal why the character was looking down but a description such as "Looks down in remorse" gives affective framing and situational context, enriching interpretation. However, depending on context, the concise description, "*looks down*" can be sufficient for interpretation. This is because descriptions can be decomposed and recomposed, meaning they are flexible enough to adjust their density and specificity depending on the amount of missing information and the user's cognitive needs, into "Literal concise" descriptions and "Detailed" descriptions.

Literal, concise descriptions engage low-level symbolic abstraction. They transmit essential situational content efficiently, minimizing interpretive ambiguity. These descriptions act as direct affordance indicators, allowing the perceiver to reconstruct the nonverbal event

mentally with a high degree of clear, easy-to-access meaning. For example, "Awkward Smile, Smiles and looks sideways" from the Audio Zoom call *Birthday Surprise*.

Detailed descriptions, in contrast, include both the immediate action and its potential affordances, for example, "Looks down in remorse" from the *Bojack Horseman* video clip. Here, the perceiver not only imagines the action but also anticipates its social or emotional implications, creating multi-layered affordances. In this way, descriptions also function as semiotic disambiguators, providing linguistic precision that clarifies the meaning of otherwise ambiguous or indeterminate cues. When perceivers encounter ambiguity; situations in which a nonverbal cue (such as a gesture, tone, or sound effect) could support multiple plausible interpretations, descriptions intervene by fixing referential meaning, thereby narrowing interpretive possibilities, and reducing uncertainty. Through elaboration, descriptions anchor the interpretation to a stable conceptual frame. This can be effective for someone who is not conceptually aware of the meanings of linguistically translated visual cues (i.e. someone who has been blind since a young age).

Descriptions also afforded confirmation - when perceivers partially interpreted a cue correctly, descriptions validated their reading and resulted in the alignment between the sender's intended meaning and the receiver's reconstructed meaning. In this way, they also afforded conceptual specificity, grounding the experience in shared linguistic categories. Thus, when a cue was ambiguous, descriptions did not merely add more information; they recalibrated the entire semiotic field, aligning perception with the intended meaning and enhancing the reliability of interpretation and further enriching it; adding layers of meaning and affective resonance even when comprehension could have been achieved without them. The participants enjoyed the extra information, for example, from the audio Zoom call example *Birthday Surprise,* in which the extra information or meaning provided, while not functionally required, enhanced subjective interpretation.

Based on this, and depending on preference, in the example of the online meeting, descriptions can immediately convey information when the BLV colleague joins. For instance, literal concise descriptions can be provided, for example, “The manager looks tired”, “Manager's posture is slouched” or "Manager has an angry frown" for directly indicating stress or fatigue.

Detailed descriptions can include, for instance, “Manager has a slumped posture and is holding his head and looking down. He has an angry frown which can indicate stress and frustration about a matter”. Another example could be “The manager’s shoulders are hunched, eyes squinting at the screen, rubbing forehead.” This can potentially add nuance and emotional context, helping the BLV individual infer that the manager is overwhelmed, allowing them to adjust how and when they speak up.

However, though descriptions served as an effective verbal-symbolic semiotic resource for clarifying or enriching meaning, they also introduced cognitive overload (Coppin, Hung, Ingino, Quevedo, Sukhai and Syed, 2024) when not carefully calibrated to user needs and

cognitive context. This was evident in the *Kung Fu Panda Shifu* video when descriptions were provided during the video and it was perceptually clashing with the dialog. Other examples shared by participants from personal experience of using Audio Description (AD) where descriptions became overstimulating and distracting descriptions as a result of the clash. This clash can also be a result of filtering redundant information from new or perceived information, and thus making added descriptions unnecessary. An example of this is seen in the *Ocean's 11* video, where AI summarised descriptions were redundant for the participants and therefore unnecessary.

## 5.5 Music

Chion (1994), cites music as a second key attribute of sound: music may be used two ways in film: empathetic (following the rhythms, tone, and emotion of the scene with which it is presented) or anempathetic (music that proceeds steadily, indifferent to the scene which it is supposedly representing).

Music acts as an affective semiotic resource, shaping the emotional and atmospheric interpretation of a scene beyond the literal ideational content. In fact, a study on sound design by (Görne, 2019), adds to this point, how further semantic and emotional communication might be achieved by means of musical structures, namely rhythm and harmony. It operates primarily through affective affordances — it evokes general emotional atmospheres (e.g., tension, calm, excitement) rather than conveying specific, propositional meaning. For example, a slow, minor-key melody might afford feelings of sadness or nostalgia (e.g. the sad music chime in the *BoJack* video clip), while an upbeat, major-key rhythm might afford excitement or joy (e.g. the harp music in the *SpongeBob Monster* video). These affordances enrich the interpretation by embedding emotional tones directly into the sensory experience of the scene. Rather than needing explicit explanation, music pre-activates emotional frameworks, making interpretation more immediate and felt. Furthermore, music was adding an additional, nonverbal layer of meaning that interacts with verbal or visual information. Even when emotional cues are present visually or verbally, music amplifies, deepens, or nuances those cues, creating multimodal emotional convergence.

Thus, music does not simply accompany a scene; it actively constructs affective meaning spaces, guiding emotional interpretation and heightening the depth of experiential engagement. This means that while music is highly effective for modulating affective tone but less reliable for ideational specificity. For example, suspenseful music may afford a sense of unease or alertness as observed from the *Teen Titans Birthday* video clip (scene 1), but it does not disambiguate why the scene is tense, who is involved, or what is actually happening. In other words, music was acting as a spatial and topological cue as an alternative to body language for perceiving the mood of a person. For instance, referring to the same example from Lee, Sukhai and Coppin (2022)'s paper of the manager's stressed mood being inferred from their posture and tone, visual cues were conveying the stressful air and in the same way, music was conveying the suspenseful air of the scene. Based on this, in the online meeting

example, music such as a muffled low-frequency hum can potentially communicate a sense of pressure or unease and in this way, music can give an immediate emotional impression of the meeting's mood.

Additionally, music's interpretive impact is highly context-dependent and shaped by cultural or experiential familiarity. Moreover, the abstractness of music (i.e. its lack of fixed linguistic referents) means that it cannot stand alone as a precise explanatory mode. It functions best when paired with more grounded modalities, such as language (descriptions) or visuals, that provide semantic anchoring and narrative framing (Coppin et al., 2024).

# 6.0 Conclusion

This study set out to explore the accessibility of nonverbal communication for blind and low-vision (BLV) individuals in online settings, emphasizing the importance of understanding how nonverbal cues are interpreted through auditory rather than visual channels. It highlights that nonverbal cues such as gestures, facial expressions, and bodily actions carry a degree of objectivity in their ability to convey signified meaning. However, in the absence of visual access, the tone of voice emerged as a powerful, albeit limited, alternative. While tone can hint at emotional and contextual information, it often lacks precision and clarity, particularly when cues are ambiguous or unfamiliar to the listener.

The use of cinematic media as a prompting tool enabled participants to recall and reflect on their own experiences, making visible the nuances of cue interpretation that are otherwise difficult to articulate. Through this method, participants not only identified which nonverbal cues were most often missed, but also contributed to the co-design of auditory alternatives such as sound effects, verbal descriptions, and music that could serve as supplementary or substitute modes of understanding.

The findings of this research reveal that literal concise descriptions were effective in providing space for independent interpretation of linguistically described visual cues. While it reduces risks of redundancy, it can result in ambiguous and unclear interpretations for a blind person who is not conceptually familiar with associated meanings due to constricted conceptual maps of visual cues and their affordances. In that case, detailed descriptions can be effective in providing clear meanings of described visual cues. Concrete sound effects are effective for indexical interpretation and abstract sound effects are for affective affordances. However, they need to be learned for instant recognition and successfully affording clear interpretation. Lastly, music also affords emotional interpretation, which can be effective in setting the mood.

The final table of findings, which categorizes audio cues by type and their perceptual or conceptual affordances, offers a practical resource for guiding future designs in accessible communication. Ultimately, this project opens new pathways for imagining inclusive online

meeting environments that better serve the interpretive needs of BLV individuals by accounting for the often overlooked layer of nonverbal expression.

# 7.0 Future Work and Recommendations

| | | Findings | Implications |
|---|---|---|---|
| **Descriptions** | Literal, concise | Literal concise descriptions were supported because they provided space for independently interpreting the meaning of the cues.<br>They were direct indicators linguistically describing expressed visual cues<br>Risk of redundant information is less | **"She's nodding"** — conveys agreement or understanding.<br>**"He's raising his eyebrows"** — indicates surprise or curiosity.<br>**"They're clapping"** — signals approval or celebration. |
| | Detailed with affordances | Description of nonverbal visual cue and potential affordances can be effective for blind people who are conceptually unfamiliar with the associated meaning of the cues (when interpretation from literal descriptions remains ambiguous, incomplete).<br>Acts as a bridge for informing meaning or confirming interpreted meaning of described cues | **She's slowly nodding with her arms crossed, signifying she is getting angry"** — may convey reluctant agreement.<br>**"He throws his hands in the air and rolls his eyes, signifying frustration"** — expresses frustration or sarcasm.<br>**"Her smile fades and she turns away slightly meaning she is sad"** — suggests discomfort or disagreement. |
| **Sound effects** | Concrete | Perceptually immediate and specific.<br>They afford instant recognition and are effective for direct indexical interpretation, identifying clear physical actions non-linguistically.<br>Easier to learn and becoming familiar with when used for signifying actions. | **"Clapping sound"** — signifying someone clapping.<br>**"Crushing sound"** — someone gesturing crushing something.<br>**"Party music sounds"** — someone gesturing upbeat dance moves |
| | Abstract | Abstract sounds are only effective when one becomes familiar with the meanings associated with it. When the abstract sounds are repeated to an extent that the perceiver can instantly perceive it and understand the meaning,<br>Can be effective for non-linguistically conveying visually expressed reactions in online meetings; a person's questioning look, combined shock reaction, silently clapping | **"Ding"** — a participant has a realisation or an idea.<br>**"Whoosh"** — someone gestured quickly or shifted position dramatically.<br>**"Pop"** — someone have an questioning look |
| **Music** | | More conceptually specific in understanding.<br>Effective for setting a mood, enhancing feelings.<br>Becoming familiar with its affordance so it can be instantly interpreted. | **A short upbeat jingle** — signals a cheerful or successful moment, like a task being completed or a warm welcome.<br>**Slow, melancholic string tone** — conveys disappointment, reflection, or tension.<br>**Mysterious low-pitched hum** — signals confusion or an unresolved issue in the conversation. |

Table 1: Suggestive Implications for Future Work

Table 4 shows a list of findings and their implications in ICT platforms based on suggestions shared by participants from the Co-design sessions. The suggestions emerged after the participants observed the audio cues' impact on their interpretation and the affordances and limitations that surfaced from it.

This project identified a range of audio cues—descriptions, sound effects, and music—that can support blind and low-vision (BLV) individuals in interpreting nonverbal communication during online meetings. They can be applied in various situations for instance, literal descriptions can offer concise translations of visual gestures, such as nodding or clapping, helping convey meaning without excess detail. Detailed descriptions add emotional and contextual nuance, particularly useful in complex social moments, such as "nodding with her arms crossed, signifying she is getting angry". Concrete sound effects, like clapping and crushing sounds, mirror real-world actions and can signal movement or activity, while abstract sound effects—such as a "ding" or "whoosh"—can symbolically convey emotions or expressions that are otherwise unseen. Music also plays a key role by setting the emotional tone of the interaction; a cheerful jingle might indicate celebration, while a somber melody can convey tension or disappointment.

Building on the current findings, future research can expand the scope by investigating in depth of the affordances and limitations of remaining cues; concrete sound effects, detailed descriptions with affordances and music and incorporating as well as the investigated audio cues in real-life online communication settings to capture a broader range of nonverbal

interactions and observe their impact. Additionally, studying recordings from live meetings can be used to study nonverbal cue accessibility which are closer to real world perceptual experiences and explore the effectiveness of audio cues from any gaps in interpretation that surfaced based on feedback.

The flexibility of some audio cues like descriptions creates opportunities for customisation. For example, after studying the different types of descriptions and their interpretational impact on meaning making, they can be integrated as short and literal verbal cues or automated summaries like, “[Person] is nodding in agreement” or “[Person] appears hesitant” to describe gestures or facial expressions. These can be triggered manually or by AI moderation. in using different formats and can be chosen by BLV users based on their preferences.

Participants suggested situations in online meetings when concrete and abstract sound effects can be useful - for signifying actions like silent clapping and surprised or shocked reactions of people during a video call. However, they also added that these nonlinguistic audio cues can only be effective if they are learned - once they become familiarised with. An example of this is seen in *Teen Titans* Birthday video (scene 11), when a participant suggested a more familiar and instantly recognisable sound, the “ding” or “bell” sound effect as an alternative to the cash register opening sound used. Therefore, future work can involve co-designing for developing a sound language - a soundtrack of audio cues for conveying or enhancing information. Sound language  is an area that can be explored for identifying which type of concrete and abstract cues can be effective and well integrated in online meetings. Familiarity with accessibility techniques is crucial as stated by (Lopez et al., 2022) and therefore, Longitudinal studies can also be used to examine how repeated exposure to different audio representations influences learning, interpretation accuracy, and emotional resonance in digital conversations.

Based on another suggestion by a participant, developing an interactive prototype that is embedded with customizable auditory cues such as adaptive sound effects, or contextual audio descriptions and spatial audio which could help test usability and personalization in real-time. Additionally, doing a deeper cognitive semiotic analysis across diverse user experiences could further refine our understanding of how BLV individuals construct meaning from non-visual cues.

Ultimately, cross-disciplinary collaboration between accessibility researchers, media designers, and the BLV community will be essential in co-creating inclusive audio-visual experiences that go beyond compensating for visual loss and toward enriching communication for all.

# References

Anderson, D. and Kelliher, C. (2020), “Enforced remote working and the work-life interface during lockdown”, Gender in Management, Vol. 35 No. 7, pp. 677-683.

Coppin, P., Hung, P., Ingino, R., Quevedo, A.U., Sukhai, M., and Syed, A.R. (2024, March). A Study of Accessible and Inclusive Virtual and Blended Information and Communication Technologies (ICTs) for the Federal Public Service and Federally Regulated Industries in Post-COVID-19 Canada. Accessibility Standards Canada. Retrieved from https://sites.google.com/view/asc-virtual-icts-accessibility/.

Baumeister, R.F., Leary, M.R.: The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol. Bull. 117(3), 497–529 (1995)

Bazin, André. What is Cinema? Ed. Hugh Gray. Berkeley: University of California Press, 1967, 1971.

Birdwhistell, R. L. (1952). Introduction to Kinesics: (An Annotation System for Analysis of Body Motion and Gesture). Department of State, Foreign Service Institute.

Botzer, A., Shvalb, N.: Using sound feedback to help blind peo- ple navigate. In: Proceedings of the 36th European Conference on Cognitive Ergonomics, Article 23, p. 3. ACM (2018)

Borkenau, P., Mauer, N., Riemann, R., Spinath, F.M., Angleit- ner, A.: Thin slices of behavior as cues of personality and intelligence. J. Pers. Soc. Psychol. 86(4), 599–614 (2004)

Brock, M., Kristensson, P. O.: Supporting blind navigation using depth sensing and sonification. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pp. 255–258. ACM (2013)

Caldwell, D., Knox, J. S., & Martin, J. R. (Eds.). (2022). APPLIABLE LINGUISTICS AND SOCIAL SEMIOTICS: Developing Theory From Practice (1st ed.). Bloomsbury Publishing Plc. https://doi.org/10.5040/9781350109322

Casetti, F., & Leisawitz, D. (2011). Sutured Reality: Film, from Photographic to Digital. October, 138, 95–106.

Chion, M., Gorbman, C., & Chion, M. (2001). Audio-vision: Sound on screen (Nachdr.). Columbia Univ. Press.

Coppin, P. (2014). Perceptual-cognitive properties of pictures, diagrams, and sentences: Toward a science of visual information design. Doctoral dissertation, University of Toronto.

Dash, D. B. B. (2022). *Significance of Nonverbal Communication and Paralinguistic Features in Communication: A Critical Analysis*.

Ellen A. Isaacs, John C. Tang. 1994. What video can and cannot do for collaboration: A case study. Multimedia Systems 2, 2, 63-73.

Fichten, C. S., Tagalakis, V., Judd, D., Wright, J., & Amsel, R. (1992). Verbal and Nonverbal Communication Cues in Daily Conversations and Dating. The Journal of Social Psychology, 132(6), 751–769. https://doi.org/10.1080/00224545.1992.9712105

Fiorelli, L. (2016). WHAT MOVIES SHOW: REALISM, PERCEPTION AND TRUTH IN FILM.

Free Text to Speech & AI Voice Generator. (2025, May 7). ElevenLabs. https://elevenlabs.io

Görne, T. (2019). The Emotional Impact of Sound: A Short Theory of Film Sound Design. 17–2. https://doi.org/10.29007/jk8h

Galioto, G., Tinnirello, I., Croce, D., Inderst, F., Pascucci, F., Giarré, L.: Sensor fusion localization and navigation for visually impaired people. In: 2018 European Control Conference (ECC), pp. 3191–3196. IEEE (2018)

Grice (1957, p.377) Grice, H.P. "Meaning." The Philosophical Review 66.3 (1957): 377-388.

Goharrizi, Z. E. (2010). Blindness and initiating communication (Master's thesis).

Humphrey, V.F. (no date) 'Overcoming the Loss of Nonverbal Cues Encountered by the Adventitiously Blind: Reconstructing Relationships and Identity'.

Kemp, N.J., Rutter, D.R.: Social interaction in blind people: an experimental analysis. Hum. Relat. 39(3), 195–210 (1986)

Kim, H. N., & Taylor, S. (2024). Differences of people with visual disabilities in the perceived intensity of emotion inferred from speech of sighted people in online communication settings. Disability and Rehabilitation: Assistive Technology, 19(3), 633–640. https://doi.org/10.1080/17483107.2022.2114555

Kleck, R.E., Nuessle, W.: Congruence between the indicative and communicative functions of eye contact in interpersonal relations. Br. J. Soc. Clin. Psychol. 7(4), 241–246 (1968)

Knapp, M., Hall, J., Horgan, T.: Nonverbal Communication in Human Interaction, 8th edn. Wadsworth Cengage Learning, Boston (2014)

Krishna, G. (2024). Making Social Interactions Accessible for the Blind and Low-Vision People by Detecting Emotions through Vocal Tone, Facial Expressions, and Body Language.

Lee, E., & Coppin, P. (2022). How virtual work environments convey perceptual cues to foster shared intentionality during Covid-19 for blind and partially sighted employees. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44, No. 44).

Lopez, M. J., & Pauletto, S. (2009). THE DESIGN OF AN AUDIO FILM FOR THE VISUALLY IMPAIRED.

Lopez, M., Kearney, G., & Hofstädter, K. (2022). Seeing films through sound: Sound design, spatial audio, and accessibility for visually impaired audiences. British Journal of Visual Impairment, 40(2), 117–144. https://doi.org/10.1177/0264619620935935

Luebstorf, S., Allen, J. A., Eden, E., Kramer, W. S., Reiter-Palmon, R., & Lehmann-Willenbrock, N. (2023). Digging into "Zoom Fatigue": A Qualitative Exploration of Remote Work Challenges and Virtual Meeting Stressors. Merits, 3(1), 151–166. https://doi.org/10.3390/merits3010010

Mamurova, S. (2024). BEYOND WORDS: THE ROLE OF PARALINGUISTICS IN EFFECTIVE COMMUNICATION. QO'QON UNIVERSITETI XABARNOMASI, 13, 300–302. https://doi.org/10.54613/ku.v13i.1083

Naufaldi, R., Wuli Fitriati, S., & Suwandi, S. (2022). The Relation of Verbal and Non-Verbal Communication to Produce Meaning in the Movie. English Education Journal, 12(3), 364–372. https://doi.org/10.15294/eej.v12i3.60822

Park, S. Y., & Whiting, M. E. (2020). Beyond Zooming there: Understanding nonverbal interaction online.

R. Riedl: "The Consequences of Causal Thinking," in: Watzlawick (ed.): Invented Reality: How Do We Know What We Think We Know? Contributions to Constructivism, Piper 1985, 10th ed. 2016

Reed, K.M. and Allen, J.A. (2022), Suddenly Hybrid: Managing the Modern Meeting, John Wiley & Sons, Hoboken, New Jersey.

Sagheer, I., Khalid, A., & Sarwer, S. (2024). Non-Verbal Communication Through Visual Storytelling: UMBRELLA Animated Short Film. Pakistan Languages and Humanities Review, 8(1), 277–292. https://doi.org/10.47205/plhr.2024(8-I)25

Shi, L., Tomlinson, B. J., Tang, J., Cutrell, E., McDuff, D., Venolia, G., Johns, P., & Rowan, K. (2019). Accessible Video Calling: Enabling Nonvisual Perception of Visual Conversation Cues. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–22. https://doi.org/10.1145/3359233

Shinohara, K. Wobbrock, J. O.: In the shadow of misperception: assistive technology use and social interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 705–714. ACM (2011)

Souriau: “The Structure of the Cinematic Universe and the Vocabulary of Filmology” (originally published in 1951), montage/av Vol. 6 (2), 1997

Steen, M. (2013). Co-design as a process of joint inquiry and imagination. Design Issues, 29(2), 16–28. https://doi.org/10.1162/DESI_a_00242

Sutton, J., & Austin, Z. (2015). Qualitative Research: Data Collection, Analysis, and Management. The Canadian Journal of Hospital Pharmacy, 68(3), 226–231.

Qiu, S. et al. (2020) ‘Understanding visually impaired people’s experiences of social signal perception in face-to-face communication’, Universal Access in the Information Society, 19, pp. 1–18. Available at: https://doi.org/10.1007/s10209-019-00698-3.
Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and Vision Computing, 27(12), 1743–1759. https://doi.org/10.1016/j.imavis.2008.11.007

Walton, Kendall L. “Transparent Pictures: On the Nature of Photographic Realism.” Critical Inquiry 11.2 (1984): 246-277. P. 251.)

# Appendix A: Media and Audio Zoom Call Examples

**MEDIA EXAMPLES :**

**BoJack Horseman - Herb video** — 0:00 - 1:36
https://www.youtube.com/watch?v=K03Y2FP9qhI

**Avatar: The last Airbender - Sokka video** — 0:44 - 0:57
https://www.youtube.com/watch?v=5929JJPZi9Y&t=71s

**SpongeBob SquarePants -Squidward video** — 2:43 - 3:09
https://www.youtube.com/watch?v=ZCsoU3JLlZU

**SpongeBob SquarePants -Monster video** — 3:10 - 3:23
https://www.youtube.com/watch?v=ZCsoU3JLlZU

**Ocean's 11 video** — 0:00 - 2:20
https://www.youtube.com/watch?v=-p0hB3a8uag&t=2s

**Steven Universe - Work video** — 0:00 - 0:52
https://www.youtube.com/watch?v=fy9TsGOswVA

**Regular Show - Benson video** — 0:00 - 0:30
https://www.youtube.com/watch?v=x7AirZHAZ8g

**Kung Fu Panda - Shifu video** — 2:00 - 3:53
https://www.youtube.com/watch?v=w5IXKURu_O0

**Kung Fu Panda - Oogway video** — 0:00 - 1:23
https://www.youtube.com/watch?v=PSBfcpqICvY

**Teen Titans Birthday video** — 0:40 - 1:33
https://www.youtube.com/watch?v=hdZvNWomDx8&t=43s

**Teen Titans - Pancake video** — 11:30 - 12:48
https://www.youtube.com/watch?v=hdZvNWomDx8&t=43s

**Audio Zoom Call Examples:**

https://www.dropbox.com/t/jIGhJ8hUHMUUioVL