Media format matters: User engagement with audio, text and video tweets

Jamy Lia\*and Mohsen Ensafjooa

a: Department of Mechanical and Industrial Engineering, Toronto Metropolitan University, 350 Victoria St., Toronto, Ontario M5B 2K3.

Emails: jamy@torontomu.ca, mensafjoo@torontomu.ca

ORCIDs: https://orcid.org/0000-0003-2440-9719, https://orcid.org/0000-0002-2994-5142

\*: Corresponding author: jamy@torontomu.ca

Jamy Li has a Ph.D. degree in communication research (Stanford University, 2016), a Master of Applied Science degree in Industrial Engineering (University of Toronto, 2008) and a Bachelor of Applied Science degree in engineering science (University of Toronto, 2006). He was an Assistant Professor in Human-Media Interaction at the University of Twente, The Netherlands, from 2016 to 2020. He is currently an Assistant Professor of Mechanical and Industrial Engineering at Toronto Metropolitan University since 2020. His research interests include human–robot interaction and user-centred design of agent-based systems for special populations. He has published in Computers in Human Behavior, IJHCS, IJSR and several human-technology interaction conferences.

Mohsen Ensafjoo is currently pursuing a Ph.D. degree in the Mechanical and Industrial Engineering program at the Toronto Metropolitan University (TMU), Toronto, Canada, under the supervision of Dr. Jamy Li. He received a bachelor's degree in mechanical engineering from Ferdowsi University of Mashhad and a master's degree in mechatronics from Iran University of Science and Technology (IUST) in Tehran, Iran. His current field of research is human-computer interaction. He has experience in robotics and automation and is interested in researching the relationship between humans and emerging new technologies.

All authors have agreed to the submission.

The article is not currently being considered for publication by any other print or electronic journal.

Media format matters: User engagement with audio, text and video tweets

**8106 words**

A dataset of shortform audio-only tweets with video and text controls is used to analyze whether a tweet’s media format and topic influence user engagement with the tweet. Audio tweets were more engaging than text or video tweets by the same users, linking engagement to media format. Highly engaging audio tweets may have had different topics than highly engaging text or video tweets and different content characteristics than low engaging audio tweets. This suggests audio-based social media networks lead to higher user engagement than text and identifies content categories that users can profitably focus on to attract and engage followers.

# Introduction

Audio social media messages include audio messages that can be communicated on social media platforms. They represent a unique media format of growing interest to consumers and media researchers alike. As examples of their overall prevalence, sixteen percent of U.S. participants in one study sent voice messages once a day (Haas et al., 2020) while WeChat users sent 6.1 billion voice messages every *day* (Tencent, “The 2017 WeChat Data Report”, November 9, 2017, https://blog.wechat.com/2017/11/09/the-2017-wechat-data-report). How do the characteristics of individual social media messages – specifically, their topic and media format – affect people’s engagement with those messages? Past research has found that audio social media messages can be about different topics. For example, analysis of Brazilian WhatsApp audio messages found their content was mostly political (Maros et al., 2020). Media format may also be an important consideration in social media messages. For example, qualitative studies of social media platforms’ users identified reasons and strategies they used voice messages/memos as opposed to other media formats (Karapanos et al., 2016; Patel et al., 2010). A novel investigation into whether and how media format (i.e., modality), topic and other attributes of audio tweets influence public engagement with those tweets may provide insight into guidelines for audio tweet engagement and the impact of media format on selection and sharing of online content.

Public engagement data may be used by social media platforms to determine how to optimally promote platform content to “steer” the activities of their users, cf. (D. Nieborg & Poell, 2019). This “platformization”, which prioritizes the needs of the platform companies, may lead to a variety of infrastructures that modularize cultural content into different packages and that content producers will have to adapt to (D. B. Nieborg & Poell, 2018). Although the economic (e.g., market), governance (e.g., content standards) and infrastructure (e.g., content unbundling due to platforms) determinants of platformization have been explored (D. B. Nieborg & Poell, 2018), less work has focused on the specifics of the different packages used – i.e., media format. Although media format may be an aspect of content standards or software development (e.g., standards may mention or software may implement specific media formats), a media studies perspective may be uniquely situated to provide actionable insights about media format. For example, a large literature presents guidelines related to social media, such as those related to online game content creation (Fränti & Fazal, 2023), social media metrics (Peters et al., 2013), content moderation (Jiang et al., 2020) and ethical use (Bowen, 2013), but a gap exists in suggesting guidelines for highly engaging message content based on format. An understanding of what makes for engaging audio content may be used both by platform companies aiming to steer content creators and by content creators who want to navigate future “platformization” actions of platforms.

Message or information characteristics (for example, information novelty, efficacy, persuasiveness or emotionality) can affect the selection or sharing of online content and should be studied since they can be manipulated and may influence recipients’ psychological motivations for engagement (Cappella et al., 2015). Moreover, researchers can speculate on these motivations prior to more complex examinations of such processes (Cappella et al., 2015). Past analyses on the popularity of social media posts by businesses, for example, revealed that functional appeals, emotional appeals and information search cues increased engagement, perhaps because such designs tapped into recipients’ psychological motivations (Swani et al., 2017). Other work found that people do not share content that reflects negatively on themselves when sharing with multiple versus a single person (Barasch & Berger, 2014). These works, however, did not look at media format. An investigation into content type, particularly in relation to media format, provides additional insight into how message characteristics motivate the selection and sharing of online content that is relevant to audio researchers.

Audio tweets may be a particularly interesting but underexplored media format to investigate. Introduced by Twitter[[1]](#footnote-2) in June 2020, audio tweets allow users to record and share audio on the Twitter platform, which was already used to share text and visual media (i.e., videos, images and animated GIFs). Audio tweets appear as regular embedded media posts on the user’s timeline that can be clicked to play and may have accompanying text, except where the video component of the media is just the user’s static profile image. Audio tweets were launched as a test venture into audio, perhaps to dovetail on the growing popularity of podcasts and to “create a more human experience” (Patterson & Bourgoin, 2020). Although initially popular, they may have reduced in popularity because of accessibility issues (Welch, 2020). Their popularity may be related to the power of audio-only content to provide “human touch” that is realized through events such as the Covid-19 pandemic. They enable users to publicly broadcast shortform audio-only messages to large audiences, rather than longform content such as podcasts (Dowling & Miller, 2019). Audio tweets are listened to as part of a social media network’s feed, whereas podcasts may be positioned as more of a “destination” closely related to radio (Berry, 2020). Past studies have looked at podcasts’ use of social media (e.g., (Adler Berg, 2021; Bonini, 2014; Bonini et al., 2014; Ferguson & Greer, 2018; Wrather, 2016); cf. (Bonini et al., 2020)), the impacts of radio’s social media use, such as the public surfacing of opinions and emotions otherwise private (Bonini, 2014) and intimacy through podcasting (Adler Berg, 2021; Miller et al., 2022). Audio tweets present a use case of shortform audio-only messages being embedded in and broadcast as part of a social media network itself, which may result in content that is different from podcast or radio. Given that Twitter supports multiple media formats, audio tweets can also be compared and contrasted with text or video/image tweets from the same people (J. Li & Penaranda Valdivia, 2022). The current work presents a quantitative Bag-of-Words (BOW) analysis as well as a qualitative content analysis of an existing corpus of real-world audio tweets, text tweets and video/image tweets from the same users (Figure 1). The influence of media format (i.e., audio modality) and lexical attributes (length, toxicity, sentiment and content topic) on user engagement with the tweets is studied. Our results can help researchers better understand the impact of audio social media messages on public discourse and can help social media network companies identify content strategies targeted on user engagement.



Figure 1: A multimodal dataset of tweets is analyzed to explore the influence of media format and lexical features on user engagement, resulting in design implications for audio social media networks.

# Background

Media format may affect user engagement with tweets. People process content in ways specific to the format the content is processed in, which may impact their engagement, cf. (Grewal et al., 2021). The mere presence of images in Twitter posts about corporate brands positively impacted user engagement by increasing retweets by 213% and likes by 151% (Y. Li & Xie, 2020). Similarly, photos and video links in Facebook brand pages increased engagement as measured by likes (Dhaoui & Webster, 2021). Video or image tweets received higher engagement than text tweets, based on a dataset of UK radical right tweets (Sprejer et al., 2022). Similarly, analysis of four official Canadian municipal Twitter accounts found that in two of the accounts, tweets with multimedia had higher user engagement than text-only tweets; in the other two accounts, no significant difference was found (Indratmo et al., 2020). The presence of videos in tweets of six Pakistani micro-celebrities reduced online engagement versus tweets without video (Iqbal Khan & Ahmad, 2022). Given these mixed findings, we test whether audio-only tweets are more engaging versus traditional text or image/video tweets.

Research Question 1 (Media format): Does tweet media format affect engagement in the tweet?

Recent work on content length found that longer tweets were more engaging than shorter tweets (Iqbal Khan & Ahmad, 2022). The text length of posts on a Facebook retail business page was significantly related to user engagement as expressed by likes (Gkikas et al., 2022). An effect in the opposite direction, however, was found among tweets of a Turkish Women’s International Network (Semiz & Berger, 2017). No significant effect of tweet length was found in a study of corporate tweets in a range of industries (Han et al., 2019). We therefore ask whether tweet length will result in higher or lower tweet engagement.

Research Question 2 (Content Length): Does content length affect engagement in the tweet?

Past work has found that toxic tweets used for the initial or first reply in a conversation result in a longer conversation chain and thus, higher engagement than non-toxic tweets (i.e., conversation length as a measure of engagement) but that within toxic tweets, higher toxicity results in fewer replies (Salehabadi et al., 2022). However, other work has found that the more toxic the tweet, the higher the engagement measured as sharing retweets, replies or quotations (Sprejer et al., 2022). A longitudinal field experiment also found that lowering toxicity by blocking toxic content resulted in lower engagement with social media content (Beknazar-Yuzbashev et al., 2022). Thus, we test whether toxicity affects user engagement among tweets.

Research Question 3 (Toxicity): Does tweet toxicity affect engagement in the tweet?

Sentiment may also influence engagement in audio tweets. In a WhatsApp study, audio messages with over 20 shares had different linguistic features – in particular, more negative sentiment words and more negations – than audio messages shared only once (Maros et al., 2020). Tiktok news videos with negative sentiment similarly had higher audience engagement than those without negative sentiment (Cheng & Li, 2023). However, the sentiment of tweets from companies did not significantly impact engagement with those tweets (Han et al., 2019).

Research Question 4 (Sentiment): Does tweet sentiment affect engagement in the tweet?

Tweet topic may also influence engagement in audio tweets. WhatsApp audio messages with over 20 shares had different topics than audio messages shared only once (Maros et al., 2020). Moreover, ‘soft’ content was found to be more engaging than ‘hard’ content in a dataset of Pakistani micro-celebrity tweets (Iqbal Khan & Ahmad, 2022) and in a dataset of Facebook comments on Arabic-language radio stations (Al-Rawi, 2016). Different topics among tweets (Bruns et al., 2016) and Facebook posts (Joo et al., 2020) may impact user engagement. The topics of posts vary across social media platforms and affect user engagement, with some topics being popular across platforms (Aldous et al., 2019).

Research Question 5 (Topic): Does tweet topic affect engagement in the tweet?

# Materials and Methods

## Dataset

We use a Twitter dataset previously reported in (J. Li & Penaranda Valdivia, 2022) with permission from the original authors. The audio tweet dataset is a manually collected dataset containing audio, media (image or video) and text tweets for 115 users based on a keyword search for ‘audio tweets’ (note that of this writing, Google’s Search APIs do not allow searching specifically for audio tweets, which is why past work used a small manually curated dataset instead of a large dataset generated by API search queries). For each of the 115 audio tweets, the dataset has an approximate date-matched text tweet and video or image tweet from the same user; thus, there were three tweets per user. For each tweet in the original dataset, the toxicity and sentiment (valence) were calculated using Google Perspective API and Vader API, respectively, using Communalytic (<https://communalytic.com>), while the content topic codes were coded by two researchers in a collaborative coding process. The coders determined initial categories and performed an initial coding, then revised codes based on literature, then recoded, with conflicts resolved by assigning per-word codes and picking the code with the highest word count. The original dataset did not include user engagement measures, did not use the number of followers for each tweet and contained no systematic qualitative content analysis of tweet content. We therefore extended the original dataset by adding user engagement measures (number of retweets, replies, likes and favorites) by visiting the publicly accessible URLs of the tweets and conducting a qualitative content analysis of the top and bottom engagement posts. A sample of the extended dataset used in this work is shown in Table 1.

Table 1: Excerpts from high and low engagement tweets in three modalities.

|  |  |  |
| --- | --- | --- |
| **Media format** | **High engagement tweet****(toxicity, sentiment, engagement, topic)** | **Low engagement tweet****(toxicity, sentiment, engagement, topic)** |
| Audio (verbatim) | Hey this is a test.(0.06, 0, 0.30, Audio tweet functionality) | Do you remember Holly Hobbie and the other pear-headed dolls? Do you remember Holly Hobbie and the other pear-headed dolls? Remember that you are not lost and you are not faint. Do you remember Holly Coby and the other doll with the pear-head?(0.09, 0, 0, Song) |
|  | Hello, everyone. [Username] here. I would like to kick off my first ever audio tweet with a really short story, and I hope you enjoy. So it's a story of a wise man. A wise man that people have been coming to complaining about the same problems every single time…(0.02, 0, 0.17, Theatric) | I remember you as you were by Pablo Neruda, translated to English by Ms Merwin. I remember you as you were in the last autumn you were the grey beret and the steel heart in your eyes, the flames of the twilight fought on an the leaves fell in the water of your soul…(0.07, 0.54, 0, Theatric) |
|  | Hello everyone, welcome back to audio tweet number 2 and I see that my last tweet only got a couple of listens an that's fine but I'm just wondering, was that just me replaying it? Or are you guys just clicking on it to listen to it? It feels nice to be heard, and if you are listening, please tweet your favourite ice-cream flavour so I can know that you are listening to it.(0.05, 0, 0.14, Non-personal commentary) | Oh Damn, now you can finally hear my voice! What's up? Thinking about what I should go easy because I haven't had lunch yet. Maybe I should pick something and start chewing and you can guess what it is? Hmm like a mukbang video without the video? A mukbang podcast or a mukbang audio tweet? Suggest something maybe? Like comment? Suggest what I should eat in the comments. It has to be vegan though cause I'm vegan. Bye this was fun not really but cool, okay bye.(0.12, 0.36, 0.0005, Personal commentary) |
| Text (paraphrased for anonymity) | An American rapper highlighted another American rapper and a self-declared homophobic American rock musician who faces abuse allegations and you are all still streaming his album??????? You are all helping an apologist????(0.63, 0.59, 0.22, Non-personal commentary) | Watched an Indian comedy-drama at least. I loved the main actor’s performance but not the movie. An older Indian movie was better. It was good to see two Indian actors in the movie after a long while.(0.03, 0.73, 0, Non-personal commentary) |
|  | From a journalist at CBS: US Postal Service agents took a media executive into custody. It’s anticipated he will be arraigned by a US judge. @SDNYnews claims he helped defraud people with a crowdfunding campaign called “We Build the Wall”.(0.06, 0, 0.19, Current events) | You are amazing and truly an inspiration in every possible way. With love.(0.07, 0.91, 0, Salutations/Thanks) |
|  | just recognized that I need to text others back in order for my phone to not be parched.(0.15, 0, 0.11, Personal commentary) | Ohh! Thank you. Love you.(0.06, 0.81, 0, Salutations/Thanks) |
| Video/image (description or verbatim text) | Photo: Right and left photos of a young and older version of a woman.(0.24, 0, 0.19, Personal commentary) | Retweet: Music video. Text of tweet retweeted is "Out of My Mind"(0.12, 0, 0, Song) |
|  | GIF: Lady reading newspaper and laughing.(0.12, 0.49, 0.08, Non-personal commentary) | Gif: Twitter bird flying.(0.09, 0, 0, Promotion) |
|  | Photo: Menu screen of "Widgetsmith", "Pinterest", "TimePassages", "Charts", "SkyView", "PlanetFitness", "Pinterest", "Clue", "The Pattern", "Health", "Co -- Star"(0.06, 0, 0.07, Non-personal commentary) | Gif: Caption "That Friday Feeling"(0.08, 0.13, 0, Theatric) |

The rationale to focus on this dataset of early adopters was because our intent to provide guidelines for audio tweet engagement is especially relevant for new rather than established users. Given that the dataset was taken 18 months after the introduction of audio tweets (i.e., in November-December 2021) and at the time, users were still getting used to audio tweets (as demonstrated by the presence of messages about users’ first impressions of audio tweets), it contained a suitable corpus of early users of audio tweets. The tweets in the dataset are representative of their sampling frame, which is early, publicly accessible audio tweets obtained via a simple text search. We note that the decision to focus on early adopters of audio tweets means that statistical findings versus established tweet formats (e.g., text) may be due to the relative novelty of audio tweets.

The procedures involved in this research were assessed by the Research Ethics Board of [Anonymous] University (REB 2021-516) and determined to not require review, since the project relies only on information in the public domain that may not be expected to be private, falling within the federal guideline TCPS2. Informed consent was not required by an ethics committee review for this work.

## Analysis

*Quantitative Analysis*

To calculate engagement, we compute normalized and raw engagement. Raw engagement is the sum of the number of retweets/quotations, replies (i.e., comments) and likes/favorites for each tweet (i.e., total post interaction), cf. (Sprejer et al., 2022). Normalized (i.e., per-follower) engagement is the raw engagement divided by the account's followers (i.e., ‘page’ followers), similar to (Iqbal Khan & Ahmad, 2022; Siyam et al., 2020). Total post interaction is used since we are interested in summative measures rather than establishing an actor-network structure of user-affiliated interactions, e.g., (Sprejer et al., 2022).

A repeated measures analysis of covariance (ANCOVA) model is used to capture the influence of media format and each of the lexical features (the categorical and interval predictors) on tweet engagement (the interval response). All quantitative analyses are done in R version 4.2.2 and RStudio version 2023.03.0+386.

### Qualitative Content Analysis

Audio, text and video tweets were sorted into the top ten and bottom ten engaging tweets. Each of the resulting groups (e.g., high engagement audio tweets) were read or listened to by both researchers to manually note more detailed descriptions of the content of each tweet beyond its general topic code that was previously coded using a code book from the original study; these subtopics / subcodes were not grouped into a code book because of their relative uniqueness. Each researcher made notes about the resulting group in comparison with the opposite engagement group and with other modality groups to identify trends, which were then added into the analyses.

To provide a more detailed analysis of the content and topic of tweets, we manually read and performed qualitative content analysis on the tweets, focusing the comparisons made on tweets with high and low engagement scores and on the subtopics, to determine what content-related aspects of tweets influence their engagement on Twitter. A similar method was used to analyze user engagement in Facebook posts made by radio stations (Al-Rawi, 2016). For this analysis, two coders read or listened to the top ten highest- and bottom ten lowest-engagement audio, text and video/image tweets in the dataset to describe their contents and subtopics in detail and to note comparisons. The top and bottom ten tweets were selected based on there being approximately ten tweets that obtained 3% engagement (method similar to (McMillan, 2009)); we note that past work collected the top five engaging tweets (Alkhathlan et al., 2021) and top ten and bottom five engaging profiles (Veale et al., 2015), though other work has used the top 20 engaging profiles (Dubois & Gaffney, 2014). Qualitative content analysis is used to assess what characteristics or topics of tweets may make them engaging, what kinds of replies people make on highly engaging tweets and whether there are differences in characteristics and replies across media formats.

# Results

## Quantitative analysis

Research Questions 1 through 5 asked whether media format, content length, tweet toxicity, valence and topic influenced user engagement with the tweet. A repeated measures analysis of covariance (ANCOVA) on engagement per follower with media format, user type, content length, toxicity, valence and topic as independent variables[[2]](#footnote-3) found a significant effect of media format, *F(*4,340) = 2.66, *p* = 0.033, *ηp*2 =0.03. Audio tweets had higher engagement per follower than text or video/image tweets from the same user (see Figure 2). No other significant effects were found. Therefore, only Research Question 1 was answered in support of an effect.



Figure 2. Horizontal bar plot of engagement per follower by media format. Audio tweets had higher engagement per follower than text or video/image tweets from the same user.

***Exploratory qualitative content analysis***

To further explore Research Question 5, we analyzed the top ten highest- and bottom ten lowest-engagement audio, text and video/image tweets in the dataset to describe their topics and characteristics. Although the quantitative analysis did not reveal an effect of content topic (Research Question 5), mixed methods research into tweet content may generate additional findings.

### Content of high-engagement and low-engagement audio tweets

The ten highest engagement audio tweets included entertaining tweets featuring user-generated content (i.e., a song, an impression, a story and a joke), non-personal commentary (i.e., asking the audience to reply with their favorite ice cream flavor), personal commentary (i.e., a dentist trip leading to speaking with a numb mouth), three comments on the functionality of audio tweets (i.e., one saying it had features no other app had, one stating a use case and one stating it is a test) and a foreign language audio tweet. Thus, entertaining tweets featuring user-generated content and tweets on the functionality of audio tweets seemed most popular. At the time of the dataset’s capture, users were just starting to get used to audio tweets such that many high-engagement audio tweets were still on the topics of first impressions of audio tweets and their functionality. Audio tweets with high engagement also included entertaining tweets and a tweet designed to solicit listener feedback. The replies to audio tweets tended to be text tweets and tended to be either positive and encouraging in nature (e.g., about the novelty of audio tweets) or responding to the user’s question (e.g., answering vanilla bean is the commenter’s favorite ice cream flavor). Thus, audio tweets that are entertaining, that ask the audience to reply with their personal preferences or that comment on the novelty of audio tweets may lead to high engagement (see Table 2).

Table 2: Trends in high and low engagement audio, text and video/image tweets.

|  |  |  |
| --- | --- | --- |
| **Media format** | **High engagement characteristics/content topics** | **Low engagement characteristics/content topics** |
| Audio tweets | User-generated entertainment, asking audience to post preferences, comment on novelty of audio tweets | Audio problems, large asks, trivial issues |
| Text tweets | Well-known TV/film, celebrities, politics, current events in a playful or opinionated way | Expressions of thanks, lesser-known TV/film, weather |
| Video/ image tweets | Celebrities, research, videos of toys or songs, personal reveals | Current events, promotion, theatrical |

The ten lowest engagement audio tweets included entertaining tweets (i.e., two songs, one with some audio ‘popping’ and another with loud street noise, and a poem translated from Filipino), a recording of fireworks, a promotion of audio tweets (i.e., asking users to post 100 audio tweets, one a day), two personal commentaries (i.e., stating dreams related to music/traveling and wondering what to eat for lunch), one personal commentary apologizing for the user’s voice, one non-personal commentary (i.e., stating issues about removing bots on Twitter) and a salutation stating good evening. This demonstrated a wide range of content topics for low engagement audio tweets including the presence of entertaining tweets and a request to commenters (which were also found for high engagement audio tweets); however, we note that some tweets with low engagement had audio issues (i.e., songs with background noise) or much larger asks (i.e., posting an audio tweet each day), which the high engagement audio tweets did not have. Thus, audio tweets with audio problems, large asks or ‘trivial’ issues such as what to eat for lunch may lead to low engagement.

### Content of high-engagement and low-engagement text tweets

The ten highest engagement text tweets included non-personal commentary (i.e., a tweet about popular U.S. rap and rock musicians, one on political parties having multiple Twitter accounts, one on Mothership, a request to Netflix to add the TV show ‘Big Time Rush’ and an anti-Trump message), current events (i.e., welcoming a U.S. media executive being held for fraud, applauding white liberals ‘girlbossing’, or inappropriately idolizing, a female U.S. politician), personal commentary (i.e., realizing the need to text people back for the user’s phone to not be dry), salutation (i.e., congratulating a U.S. actor on his daughter’s school progress), or onomatopoeia sounds (i.e., ‘DA’). Replies to text tweets were mostly text tweets themselves. Replies to non-personal commentary and current event tweets affirmed the original tweet’s message (e.g., replying an anti-Trump comment on the anti-Trump tweet) but in some cases went against the original tweet (e.g., replying that the Justice Department is corrupt in holding a U.S. media executive for fraud), which in turn garnered replies on the reply (e.g., replying to the replier to shut up). Thus, most of the high engagement text tweets were about TV/film, celebrities (note that these categories differ from “user-generated entertainment” in that the media they focus on is commercial rather than user-generated), politics or current events in an opinionated (e.g., comment on ‘girlbossing’) or playful (e.g., comment on a U.S. actor’s daughter) manner.

The ten lowest engagement text tweets included six salutation/thanks tweets (i.e., thanks for posting a tweet), a current event (i.e., report on the weather), personal commentary (i.e., missing a dog), non-personal commentary (i.e., stating that a link another user posted was out of date, commenting on an Indian comedy-drama film). Engagement with thanking tweets may be low because thanks can come at the end of a conversation or interaction, thereby not eliciting further discourse. Engagement may also be low for a text tweet that mentioned an Indian comedy-drama film because it may be lesser known than mainstream media mentioned in high engagement tweets. Thus, most of the low engagement tweets are expressions of thanks that the user posts on their timeline, lesser-known media or the current weather.

### Content of high-engagement and low-engagement video/image tweets

The ten highest engagement video/image tweets included six non-personal commentary (i.e., an animated GIF of a woman reading a newspaper and laughing, a screenshot of Pinterest panels and musician Liam Payne memes, a link to an article on bereavement with a thumbnail of the journal Aging and Mental Health, a video of a U.S. actor and his daughters, an image of butter cookies asking which shape is the best, a video of toy figurines on a shelf with a Kermit the Frog Pez dispenser on the floor behind the shelf), a personal commentary (i.e., side-by-side photos of a woman who just turned 18 and her younger self), a salutation/thanks screenshot (i.e., expressing thanks to a screenshotted comment about the user’s professionalism at work), a theatrical post (i.e., an animated GIF of a woman drawing a clown’s face on her face as a reflection on the user’s choice not to graduate early) and a song (i.e., a video of an orchestra playing a foreign language song). Replies included commenting on butter cookie shapes, on the video of toy figurines and Pez and birthday wishes for the user who just turned 18. Thus, high engagement video/image tweets are associated with commentary on celebrities or research, custom video of toys or songs or personal reveals.

The ten lowest engagement video/image tweets included current events (i.e., a link to an article on Twitter Blue with a thumbnail of Twitter Blue, a link to an article with an image of stick figures giving money to a media platform), promotion (i.e., an advertisement for a Friday event on Zoom, an image of a five-star review for an advisor with text stating appreciation, an animated GIF of Twitter birds flying with text about coach Larry Bird), non-personal commentary (i.e., an animated GIF stating ‘You are amazing’), theatrical content (i.e., an animated GIF of The Office characters dancing with a caption stating ‘That Friday Feeling’, an animated GIF of a person rolling their middle finger up with text stating the user is turning 30 soon), a retweeted music video and an image of another user’s foreign language text. Thus, low engagement video tweets have content that is current events (perhaps because current events do not suit the modality of a Twitter video), promotional or theatrical (perhaps because the events or messages are of low interest to others).

### Content length, toxicity and sentiment of high-engagement and low-engagement audio, text and video/image tweets

Although not our focus, we also looked at content length, toxicity and sentiment between the top and bottom engaging audio, text and video/image tweets to further explore Research Questions 2, 3 and 4 despite the null quantitative results. Although content length and sentiment did not seem to differ between high and low engaging tweets, highly engaging text tweets seemed to have more potentially toxic content on politically or culturally polarizing issues than low engaging text tweets, though this pattern was not found in audio tweets.

# Discussion

## Summary of Results

Mixed methods analysis of a pre-existing dataset of audio, text and video/image tweets found that audio tweets had higher engagement per follower than did text or video/image tweets by the same user, supporting the idea that user engagement is linked to media format. Qualitative content analysis revealed that high engagement audio tweets may have had different content topics than high engagement text and video tweets, while high engagement audio tweets had somewhat similar topics but different characteristics than low engagement audio tweets. Given these differences, the media format, topic and characteristics of a tweet may affect engagement in the tweet.

## Design Implications

Table 3 shows how the current work’s findings translate into design guidelines for an audio-focused social media network. Audio tweets had higher engagement than both text and video controls from the same user. This suggests that low uptake on audio tweeting is not due to limitations in the impact made by audio tweets upon a user’s followers, but instead due to the usability of the feature itself. Audio tweets’ low uptake is more likely due to lack of awareness about the feature or high effort costs of recording an audio tweet, since the impact they have on an audience is higher than in other tweet modalities.

Table 3: Design implications for audio social media networks based on current work’s findings.

|  |  |
| --- | --- |
| **Finding****[Research Question]** | **Design Implication** |
| Audio tweets were more engaging than text or video/image tweets.[Research Question 1] | Audio-based social media networks have higher user engagement than text-based networks. |
| High engagement audio tweets may have had different topics than high engagement text or video tweets.[Research Question 5] | Audio-based social media networks and their users may profitably focus their platforms and posts on content topics of high engagement among audio tweets. |
| High engagement audio tweets had different characteristics but not different topics than low engagement audio tweets.[Research Question 5] | Designers of audio-based social media networks can give users tips or metrics to encourage users to post audio tweets with characteristics akin to high engagement. |

Qualitative analysis found that highly engaging audio tweets were not always on the same topics as highly engaging text or video tweets. Highly engaging audio tweets included user-generated entertainment (i.e., light-hearted or performative topics) and commentary on audio tweet functionality (this pattern was not seen in highly engaging text tweets), whereas highly engaging text tweets included politics and current events (which were not seen in highly engaging audio tweets). Thus, social media platforms that assess user engagement in “test” social media posts to optimize user acquisition profits (Van Dijck et al., 2018), cf. (D. Nieborg & Poell, 2019) must carefully consider and segregate the modality they use to determine what topic is suitable for that modality. Since audio-based social media networks may profitably focus their platforms on content topics of high engagement among audio tweets, such as user-generated entertainment and commentary, cultural production of audio social media may be dependent on and limited by each platform’s practices (D. Nieborg & Poell, 2019). Audio-rich platforms such as WeChat add value to users’ lives as demonstrated by their prevalence but may also serve as tools for “cultural governance” (de Kloet et al., 2019).

Highly engaging audio tweets had different characteristics (e.g., audio production quality) than low engaging audio tweets. This may mean that audio social media networks can profitably provide metrics or guidelines on audio recordings to its users to help them post high-quality recordings, which existing networks such as Twitter do not prominently provide. However, platforms’ biased steering of online interaction (i.e., “platformization”, to attract higher engagement for instance) may compromise user autonomy in the process of cultural production (D. Nieborg & Poell, 2019).

## Implication for Theory

Highly engaging audio tweets contained different content characteristics than low engaging audio tweets and contained different topics than highly engaging text and video tweets. This demonstrates that different types of content not only distinguish engaging from non-engaging tweets, but also that the content topics that are engaging in one modality may not be engaging for another modality. This is in line with past work that suggests voice has more negative sentiment than text (J. Li & Penaranda Valdivia, 2022), which may mean that people prefer delivering negative comments verbally as opposed to via text. The current work finds that media format may also influence the content topic of the post in addition to its sentiment. Audio-based social media networks may do well to focus their platforms on content topics that are popular for audio messages.

The topics of engaging text tweets appeared to differ from the topics of non-engaging text tweets; however, this trend was not as apparent among engaging versus non-engaging audio tweets. Because of their relative novelty, users may not be as familiar with what makes an engaging audio tweet, which can in turn cause greater variability in the distribution of content types among non-engaging audio tweets. Entertaining tweets were among the non-engaging audio tweets, for instance, due to low production value. Early-stage audio social media networks may need time for users to get accustomed to production methods to create successful audio posts.

## Limitations and Future Work

The findings obtained here were from a small manually curated dataset of audio, text and video tweets and may be specific to the dataset used; thus, our findings are speculations prior to more complex evaluations (Cappella et al., 2015). The generalizability of our findings to larger datasets of audio tweets and to the content in other audio media platforms is therefore an important area of further research. Similarly, this work explored only a few potential determinants of user engagement in audio tweets; in particular, further work can look at the impact of personal characteristics such as information efficacy or confidence in whether or not a user decides to engage with a tweet (Warner et al., 2012). Although we found significant differences in engagement between audio and text tweets, such differences may have been attributable to parameters that varied in our dataset, which could be clarified by a future experiment that controls for such parameters to present “fairer” comparisons. Audio and visual media are also not exclusive options, since radio stations can use Instagram or visualizations for community building, for instance (Berry, 2013; Ferguson & Greer, 2018); additional work can look at how different media types are used in tandem for community building. The qualitative analysis could also benefit from a more in-depth look at longitudinal impact beyond the introduction of audio tweets, such as the transition to “X” and resulting functionality or direction changes or how an especially popular audio tweet may impact society (cf. (O’Kelly, 2014)), from additional analysis about how a specific community such as that formed by a podcast engages users (Pavelko & Myrick, 2020) or from addressing the influence of gender and other voice features in the perception of audio content. Although the current work speculates on what content topics are engaging for audio tweets, future work exploring actual audio tweets about less popular topics such as current events may reveal insights about how to support the development of audio on those topics too.

The accuracy of perceiving a user’s intentions through social media may also be impacted by media format. Past work suggests that people may not accurately perceive social media users’ intentions when reading their social media posts as text (Brady et al., 2023) but did not explore whether the accuracy of people’s perceptions may be improved through audio or video formats. People who listened to a disagreeing political viewpoint gave higher humanization ratings than people who read text of the same viewpoint (Schroeder et al., 2017) – suggesting that voice may be more conducive to social understanding or connection than text. The potential for audio media to improve intention perception may be an interesting area for future research, perhaps with a dataset that controls for topic using propensity score matching.

# Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant number RGPIN-2021-03139.

# Declaration of Interest Statement

The authors declare that there is no conflict of interest.

# References

Adler Berg, F. S. (2021). The value of authenticity and intimacy: A case study of the Danish independent podcast Fries before Guys’ utilization of Instagram. *Radio Journal: International Studies in Broadcast & Audio Media*, *19*(1), 155–173. https://doi.org/10.1386/rjao\_00039\_1

Aldous, K. K., An, J., & Jansen, B. J. (2019). View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 47–57. https://doi.org/10.1609/icwsm.v13i01.3208

Alkhathlan, M., Tlachac, M. L., Harrison, L., & Rundensteiner, E. (2021). “Honestly I Never Really Thought About Adding a Description”: Why Highly Engaged Tweets Are Inaccessible. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Eds.), *Human-Computer Interaction – INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30 – September 3, 2021, Proceedings, Part I* (Vol. 12932, pp. 373–395). Springer International Publishing. https://doi.org/10.1007/978-3-030-85623-6

Al-Rawi, A. (2016). Understanding the Social Media Audiences of Radio Stations. *Journal of Radio & Audio Media*, *23*(1), 50–67. https://doi.org/10.1080/19376529.2016.1155298

Barasch, A., & Berger, J. (2014). Broadcasting and Narrowcasting: How Audience Size Affects What People Share. *Journal of Marketing Research*, *51*(3), 286–299. https://doi.org/10.1509/jmr.13.0238

Beknazar-Yuzbashev, G., Jiménez Durán, R., McCrosky, J., & Stalinski, M. (2022). *Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment* (SSRN Scholarly Paper 4307346). https://doi.org/10.2139/ssrn.4307346

Berry, R. (2013). Radio with pictures: Radio visualization in BBC national radio. *Radio Journal: International Studies in Broadcast & Audio Media*, *11*(2), 169–184. https://doi.org/10.1386/rjao.11.2.169\_1

Berry, R. (2020). Radio, music, podcasts – BBC Sounds: Public service radio and podcasts in a platform world. *Radio Journal: International Studies in Broadcast & Audio Media*, *18*(1), 63–78. https://doi.org/10.1386/rjao\_00016\_1

Bonini, T. (2014). Doing radio in the age of Facebook. *Radio Journal: International Studies in Broadcast & Audio Media*, *12*(1/2), 73–87. https://doi.org/10.1386/rjao.12.1-2.73\_1

Bonini, T., Fesneau, E., Perez, J. I. G., Luthje, C., Jedrzejewski, S., Pedroia, A., Rohn, U., Sellas, T., Starkey, G., & Stiernstedt, F. (2014). Radio formats and social media use in Europe—28 case studies of public service practice. *Radio Journal: International Studies in Broadcast & Audio Media*, *12*(1/2), 89–107. https://doi.org/10.1386/rjao.12.1-2.89\_1

Bonini, T., Monclús, B., & Scifo, S. (2020). Radio as a social media. *Radio Journal: International Studies in Broadcast & Audio Media*, *18*(1), 5–12. https://doi.org/10.1386/rjao\_00012\_2

Bowen, S. A. (2013). Using Classic Social Media Cases to Distill Ethical Guidelines for Digital Engagement. *Journal of Mass Media Ethics*, *28*(2), 119–133. https://doi.org/10.1080/08900523.2013.793523

Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., & Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour*, *7*(6), Article 6. https://doi.org/10.1038/s41562-023-01582-0

Bruns, A., Moon, B., Paul, A., & Münch, F. (2016). Towards a typology of hashtag publics: A large-scale comparative study of user engagement across trending topics. *Communication Research and Practice*, *2*(1), 20–46. https://doi.org/10.1080/22041451.2016.1155328

Cappella, J. N., Kim, H. S., & Albarracín, D. (2015). Selection and Transmission Processes for Information in the Emerging Media Environment: Psychological Motives and Message Characteristics. *Media Psychology*, *18*(3), 396–424. https://doi.org/10.1080/15213269.2014.941112

Cheng, Z., & Li, Y. (2023). Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos. *Social Science Computer Review*, 089443932311786. https://doi.org/10.1177/08944393231178603

de Kloet, J., Poell, T., Guohua, Z., & Yiu Fai, C. (2019). The platformization of Chinese Society: Infrastructure, governance, and practice. *Chinese Journal of Communication*, *12*(3), 249–256. https://doi.org/10.1080/17544750.2019.1644008

Dhaoui, C., & Webster, C. M. (2021). Brand and consumer engagement behaviors on Facebook brand pages: Let’s have a (positive) conversation. *International Journal of Research in Marketing*, *38*(1), 155–175. https://doi.org/10.1016/j.ijresmar.2020.06.005

Dowling, D. O., & Miller, K. J. (2019). Immersive Audio Storytelling: Podcasting and Serial Documentary in the Digital Publishing Industry. *Journal of Radio & Audio Media*, *26*(1), 167–184. https://doi.org/10.1080/19376529.2018.1509218

Dubois, E., & Gaffney, D. (2014). The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter. *American Behavioral Scientist*, *58*(10), 1260–1277. https://doi.org/10.1177/0002764214527088

Ferguson, D. A., & Greer, C. F. (2018). Visualizing a Non-Visual Medium through Social Media: The Semiotics of Radio Station Posts on Instagram. *Journal of Radio & Audio Media*, *25*(1), 126–141. https://doi.org/10.1080/19376529.2017.1385617

Fränti, P., & Fazal, N. (2023). Design Principles for Content Creation in Location-Based Games. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *19*(5s), 1–30. https://doi.org/10.1145/3583689

Gkikas, D. C., Tzafilkou, K., Theodoridis, P. K., Garmpis, A., & Gkikas, M. C. (2022). How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. *International Journal of Information Management Data Insights*, *2*(1), 100067. https://doi.org/10.1016/j.jjimei.2022.100067

Grewal, R., Gupta, S., & Hamilton, R. (2021). Marketing Insights from Multimedia Data: Text, Image, Audio, and Video. *Journal of Marketing Research*, *58*(6), 1025–1033. https://doi.org/10.1177/00222437211054601

Haas, G., Gugenheimer, J., Rixen, J. O., Schaub, F., & Rukzio, E. (2020). “They Like to Hear My Voice”: Exploring Usage Behavior in Speech-Based Mobile Instant Messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–10). Association for Computing Machinery. http://doi.org/10.1145/3379503.3403561

Han, X., Gu, X., & Peng, S. (2019). Analysis of Tweet Form’s effect on users’ engagement on Twitter. *Cogent Business & Management*, *6*(1), 1564168. https://doi.org/10.1080/23311975.2018.1564168

Indratmo, Zhao, M., & Buro, K. (2020). Comparisons Between Text-Only and Multimedia Tweets on User Engagement. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3825–3831. https://doi.org/10.1109/SMC42975.2020.9283438

Iqbal Khan, S., & Ahmad, B. (2022). Tweet so good that they can’t ignore you! Suggesting posting strategies to micro-celebrities for online engagement. *Online Information Review*, *46*(2), 319–336. https://doi.org/10.1108/OIR-08-2020-0334

Jiang, J. “Aaron,” Middler, S., Brubaker, J. R., & Fiesler, C. (2020). Characterizing Community Guidelines on Social Media Platforms. *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*, 287–291. https://doi.org/10.1145/3406865.3418312

Joo, S., Lu, K., & Lee, T. (2020). Analysis of content topics, user engagement and library factors in public library social media based on text mining. *Online Information Review*, *44*(1), 258–277. https://doi.org/10.1108/OIR-11-2018-0345

Karapanos, E., Teixeira, P., & Gouveia, R. (2016). Need fulfillment and experiences on social media: A case on Facebook and WhatsApp. *Computers in Human Behavior*, *55*, 888–897. https://doi.org/10.1016/j.chb.2015.10.015

Li, J., & Penaranda Valdivia, K. (2022). Does Media Format Matter? Investigating the Toxicity, Sentiment and Topic of Audio Versus Text Social Media Messages. *Proceedings of the 10th International Conference on Human-Agent Interaction*, 13–22. https://doi.org/10.1145/3527188.3561927

Li, Y., & Xie, Y. (2020). Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, *57*(1), 1–19. https://doi.org/10.1177/0022243719881113

Maros, A., Almeida, J., Benevenuto, F., & Vasconcelos, M. (2020). Analyzing the Use of Audio Messages in WhatsApp Groups. *Proceedings of The Web Conference 2020*, 3005–3011. https://doi.org/10.1145/3366423.3380070

McMillan, W. J. (2009). Finding a Method to Analyze Qualitative Data: Using a Study of Conceptual Learning. *Journal of Dental Education*, *73*(1), 53–64. https://doi.org/10.1002/j.0022-0337.2009.73.1.tb04639.x

Miller, K. J., Fox, K., & Dowling, D. O. (2022). From Black Lives Matter to COVID-19: Daily news podcasts and the reinvention of audio reporting. *Radio Journal: International Studies in Broadcast & Audio Media*, *20*(2), 131–152. https://doi.org/10.1386/rjao\_00063\_1

Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, *20*(11), 4275–4292. https://doi.org/10.1177/1461444818769694

Nieborg, D., & Poell, T. (2019). The Platformization of Making Media. In M. Prenger & M. Deuze (Eds.), *Making Media* (1st ed., pp. 85–96). Amsterdam University Press. https://doi.org/10.1017/9789048540150.006

O’Kelly, L. (2014). War of the Worlds to Social Media: Mediated Communication in Times of Crisis. *Journal of Radio & Audio Media*, *21*(2), 321–323. https://doi.org/10.1080/19376529.2014.950161

Patel, N., Chittamuru, D., Jain, A., Dave, P., & Parikh, T. S. (2010). Avaaj Otalo: A field study of an interactive voice forum for small farmers in rural India. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI ’10*, 733. https://doi.org/10.1145/1753326.1753434

Patterson, M., & Bourgoin, R. (2020, June 17). *Your Tweet, your voice*. Twitter Blog. https://blog.twitter.com/en\_us/topics/product/2020/your-tweet-your-voice

Pavelko, R. L., & Myrick, J. G. (2020). Muderinos and Media Effects: How the My Favorite Murder Podcast and its Social Media Community May Promote Well-being in Audiences with Mental Illness. *Journal of Radio & Audio Media*, *27*(1), 151–169. https://doi.org/10.1080/19376529.2019.1638925

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social Media Metrics—A Framework and Guidelines for Managing Social Media. *Journal of Interactive Marketing*, *27*(4), 281–298. https://doi.org/10.1016/j.intmar.2013.09.007

Salehabadi, N., Groggel, A., Singhal, M., Roy, S. S., & Nilizadeh, S. (2022). *User Engagement and the Toxicity of Tweets* (arXiv:2211.03856). arXiv. https://doi.org/10.48550/arXiv.2211.03856

Schroeder, J., Kardas, M., & Epley, N. (2017). The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement. *Psychological Science*, *28*(12), 1745–1762. https://doi.org/10.1177/0956797617713798

Semiz, G., & Berger, P. D. (2017). DETERMINING THE FACTORS THAT DRIVE TWITTER ENGAGEMENT-RATES. *Archives of Business Research*, *5*(2), Article 2. https://doi.org/10.14738/abr.52.2700

Siyam, N., Alqaryouti, O., & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society*, *60*, 101211. https://doi.org/10.1016/j.techsoc.2019.101211

Sprejer, L., Margetts, H., Oliveira, K., O’Sullivan, D. J. P., & Vidgen, B. (2022). An actor-based approach to understanding radical right viral tweets in the UK. *Journal of Policing, Intelligence and Counter Terrorism*, 1–19. https://doi.org/10.1080/18335330.2022.2086440

Swani, K., Milne, G. R., Brown, B. P., Assaf, A. G., & Donthu, N. (2017). What messages to post? Evaluating the popularity of social media communications in business versus consumer markets. *Industrial Marketing Management*, *62*, 77–87. https://doi.org/10.1016/j.indmarman.2016.07.006

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.

Veale, H. J., Sacks-Davis, R., Weaver, E. R., Pedrana, A. E., Stoové, M. A., & Hellard, M. E. (2015). The use of social networking platforms for sexual health promotion: Identifying key strategies for successful user engagement. *BMC Public Health*, *15*(1), 85. https://doi.org/10.1186/s12889-015-1396-z

Warner, B. R., McGowen, S. T., & Hawthorne, J. (2012). Limbaugh’s Social Media Nightmare: Facebook and Twitter as Spaces for Political Action. *Journal of Radio & Audio Media*, *19*(2), 257–275. https://doi.org/10.1080/19376529.2012.722479

Welch, C. (2020, June 17). *Twitter starts rolling out audio tweets on iOS*. The Verge. https://www.theverge.com/2020/6/17/21294481/twitter-audio-tweets-now-available-iphone-ios

Wrather, K. (2016). Making “Maximum Fun” for fans: Examining podcast listener participation online. *Radio Journal: International Studies in Broadcast & Audio Media*, *14*(1), 43–63. https://doi.org/10.1386/rjao.14.1.43\_1

1. We note that Twitter was rebranded as “X” in July 2023. As the dataset was collected when the platform was still called Twitter, we use that name in this work. [↑](#footnote-ref-2)
2. modeled in R as “aov(engagementPerFollower ~ mediaFormat + userType + words + toxicity + sentiment + topic + Error( id/ (mediaFormat + userType + words + toxicity + sentiment + topic) ), data = l)” [↑](#footnote-ref-3)