Audio-as-Data Tools: Replicating Computational Data Processing

Lukito, Josephine; Greenfield, Jason; Yang, Yunkang; Dahlke, Ross; Brown, Megan A.; Lewis, Rebecca; Chen, Bin

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Lukito, J., Greenfield, J., Yang, Y., Dahlke, R., Brown, M. A., Lewis, R., Chen, B. (2024). Audio-as-Data Tools: Replicating Computational Data Processing. *Media and Communication*, *12*. https://doi.org/10.17645/mac.7851

Media and Communication
2024 • Volume 12 • Article 7851
https://doi.org/10.17645/mac.7851

cogitatio

ARTICLE

Open Access Journal

Audio-as-Data Tools: Replicating Computational Data Processing

Josephine Lukito [1], Jason Greenfield [2], Yunkang Yang [3], Ross Dahlke [4], Megan A. Brown [5], Rebecca Lewis [4], and Bin Chen [1,6]

[1] School of Journalism and Media, University of Texas at Austin, USA
[2] Center for Social Media and Politics, New York University, USA
[3] Department of Communication & Journalism, Texas A&M University, USA
[4] Department of Communication, Stanford University, USA
[5] School of Information, University of Michigan, USA
[6] Journalism and Media Studies Centre, University of Hong Kong

Correspondence: Josephine Lukito (jlukito@utexas.edu)

Submitted: 16 November 2023 **Accepted:** 22 February 2024 **Published:** 6 May 2024

Issue: This article is part of the issue "Reproducibility and Replicability in Communication Research" edited by Johannes Breuer (GESIS—Leibniz Institute for the Social Sciences / Center for Advanced Internet Studies) and Mario Haim (LMU Munich), fully open access at https://doi.org/10.17645/mac.i429

Abstract

The rise of audio-as-data in social science research accentuates a fundamental challenge: establishing reproducible and reliable methodologies to guide this emerging area of study. In this study, we focus on the reproducibility of audio-as-data preparation methods in computational communication research and evaluate the accuracy of popular audio-as-data tools. We analyze automated transcription and computational phonology tools applied to 200 episodes of conservative talk shows hosted by Rush Limbaugh and Alex Jones. Our findings reveal that the tools we tested are highly accurate. However, despite different transcription and audio signal processing tools yield similar results, subtle yet significant variations could impact the findings' reproducibility. Specifically, we find that discrepancies in automated transcriptions and auditory features such as pitch and intensity underscore the need for meticulous reproduction of data preparation procedures. These insights into the variability introduced by different tools stress the importance of detailed methodological reporting and consistent processing techniques to ensure the replicability of research outcomes. Our study contributes to the broader discourse on replicability and reproducibility by highlighting the nuances of audio data preparation and advocating for more transparent and standardized practices in this area.

Keywords

audio-as-data; computational methods; conservative talk shows; data processing; reproduction; talk radio

1. Audio-as-Data Tools: Reproducing Computational Data Processing

Like other disciplines, communication and media researchers have increasingly been concerned with the replicability of the field's research (Benoit & Holbert, 2008; McEwan et al., 2018). Replicability and reproducibility are critical for research: Without them, it is unclear whether a particular finding is a consequence of nuanced research decisions or an actual finding. However, methodological obfuscation of data collection, preparation, and analysis (intentionally or otherwise) continues to plague replication and reproduction efforts. Open science efforts—particularly those tailored to our field—provide new avenues for producing more empirically grounded research (Dienlin et al., 2020), but blind spots remain.

One blind spot of interest to us, specifically in computational communication research, is data preparation: the steps for cleaning and wrangling the data for analysis. While challenges to reproducing data collection methods and sharing data persist (for more, see Van Atteveldt et al., 2019), data preparation is often glossed over or subsumed as part of the data analysis process, particularly if researchers are relying on computers (Plesser, 2018). However, replicating and reproducing data preparation processes is critical as the results of two studies may vary because of data cleaning, even when holding the analysis or collection strategies constant.

This study explores reproducing data preparation practices more concretely by focusing on audio-as-data tools for data preparation. We chose this type of data for three reasons. First, there is increasing scholarly interest in audio data (likely driven by the growing popularity of digital audiovisual content), particularly spoken language. Second, researchers often transform audio-as-data into other data forms as a part of the data preparation process, such as when researchers turn spoken language into a transcription for text-as-data approaches. Third, there has yet to be an empirical study that compares whether different audio-as-data processing tools produce similar results.

To assess whether results from audio-as-data processing tools will reproduce, we consider two important data preparation practices for audio data: automated transcription and audio signal processing for computational phonology. Using two datasets of right-wing talk shows, we compare the results of four tools, two for transcription and two for detecting pitch and intensity. Our results find that, while these tools produce similar results, the nuances of each tool produce subtle differences that highlight why replication and reproduction studies must account for variations in data preparation.

2. Literature Review

2.1. Conceptualizing Audio-as-Data

We define "audio-as-data" as approaches for computationally processing and analyzing auditory communication (including, but not limited to, music, speech, and noise) to address important questions in political and social life. The field of communication is currently dominated by the "text-as-data" approach, often because of convenience and size (Lukito, Brown, et al., 2023). Studies involving multimedia content such as radio, television, or podcasts often deal with audio data by textual transcription, during which the auditory features (e.g., pitch, loudness, tone) are lost (Dietrich et al., 2019; Knox & Lucas, 2021).

Despite audio data becoming more prominent in digital spaces, there are relatively few audio-as-data studies (Piñeiro-Otero & Pedrero-Esteban, 2022). It is not just the content of speech but its delivery that informs us; auditory cues in political speech, for instance, can subtly yet significantly indicate emotion and stance and change opinions (Dietrich et al., 2019; Klofstad et al., 2012; Knox & Lucas, 2021). It is worth noting that none of the above three studies used any additional software to validate the output of audio features produced by one software: Dietrich et al. (2019) and Klofstad et al. (2012) both used praat only to measure pitch (Jadoul et al., 2018), whereas Knox and Lucas (2021) used the "communication" R library to produce audio features (Lucas, 2022).

Spoken language, as a type of audio-as-data, can processed with two approaches: a reductive approach and an additive approach (Lukito, 2023). Reductive approaches remove information from a dataset because that information is irrelevant to a particular project or research question. In the context of audio data, the most common reductive approach transforms audio data into text, removing auditory cues and facilitating text-as-data approaches through automated transcription. Automated transcriptions have the advantage of speed and scale: Automated transcriptions take a fraction of the time compared to manual transcription. One common concern is accuracy: Older automated transcription tools often made mistakes, and manual tools were necessary to correct automated transcription (e.g., Luz et al., 2008). However, the ubiquity of digital video content has motivated demand for automated transcription, resulting in considerable improvements over the last decade, both in terms of accuracy (Bokhove & Downey, 2018) and in the variety of languages considered (Wisniewski et al., 2020).

Whereas reductive approaches remove unnecessary (in the context of a study) information, additive approaches involve highlighting or annotating information so that these features can be more easily studied. Additive approaches enrich our understanding by highlighting auditory features, employing audio signal processing for both speech and music to identify characteristics like pitch and intensity, two of the most popular auditory features studied (Gold et al., 2011; Purwins et al., 2019). Pitch refers to the frequency of the wavelength of a sound. Higher-pitched sounds tend to be shriller, with more frequent oscillations. By contrast, lower-pitched sounds are deeper. To use a musical example, sopranos are higher-pitched, and baritones are lower-pitched. Pitch algorithms have advanced the study of accents and vocal nuances like sarcasm (Iosad, 2015; Larrouy-Maestri et al., 2023). Previous studies have found that pitch can impact a political candidate's electability; Klofstad (2016), for example, found that candidates perceived to have a lower pitch generally received more votes than their higher-pitched opponents, though this may vary by gender.

Whereas pitch is described concerning highness and lowness, intensity refers to the number of sound waves passing through an area per second. Intensity and loudness are related, as a more intense sound will be perceived as louder by the human ear (for this reason, intensity and loudness measures are often highly correlated; this was also the case for our study). Intensity measures have been applied to emotional analysis and acoustic engineering (Chen et al., 2012; Indrayani et al., 2020; Larsen & Aarts, 2005). In other auditory analyses, auditory attributes such as duration (how long an auditory note is held) and timbre (distinctions between two instruments or two voices) are also considered. However, these are studied more regarding music (e.g., Krumhansl & Iverson, 1992) rather than spoken language (e.g., Dietrich et al., 2019).

Despite their growing use in disciplines such as computer engineering and linguistics, these tools have comparatively few applications in media and communication research (for an exception, see Shah et al.,

2023). This scholarship may be scant because of perceived applicability, as few studies have shown how researchers can leverage these tools to study media. However, we hope these audio-as-data methods become more accessible to our field. As communication and media scholars have long studied audio media (e.g., Christenson & Lindlof, 1983; Spinelli & Dann, 2019), it stands to reason that adopting these methods will become more widespread in the literature. If so, communication researchers must compare these tools to understand how computational phonology and audio signal processing tools should be applied.

2.2. Replicable and Reproducible Audio Data Preparation

Over the past few decades, social scientists have raised concerns about the replicability and reproducibility of research. While related, these two terms are conceptually distinct (National Academies of Sciences, Engineering, and Medicine, 2019), so it is important to define this terminology. The National Academies of Sciences, Engineering, and Medicine (2019) defines reproducibility as using the same input data and data processing and producing similar results—They describe this as "computational reproducibility" (p. 6). In contrast, replication refers to finding the same results from two separate data collections. Other definitions are more ambiguous; for example, Nosek and Errington (2020) define replication as "a study for which any outcome would be considered diagnostic evidence about a claim from prior research" (p. 2). Adding further confusion, other researchers have presented conflicting definitions; for example, Plesser (2018) defined replicability as when two different teams apply the same research design and produce similar results, whereas reproducibility refers to different research teams applying different research designs yet producing similar results (Plesser also defines "repeatability" as being able to repeat one's own research and produce similar results).

One gap in this literature is the reductive treatment of the research design, which often includes multiple steps such as data collection, data processing, and data analysis; however, collection and analysis are often emphasized compared to data processing. For example, Plesser's distinction between research repeated by the same team versus research replicated or reproduced by a different team emphasizes differences in data collection (e.g., location of the research team). Similarly, many definitions of replicability and reproducibility focus on the outcome of the research (Howell, 2020; National Academies of Sciences, Engineering, and Medicine, 2019; Nosek & Errington, 2020). While it is important to define these concepts, focusing solely on definition misses a key issue with research on replicability and reproducibility: What part of any given research process becomes unreplicable or unreproducible? Furthermore, what does replicable and reproducible mean in data preparation? Suppose a team conducts a "reproduction study," but uses different software to conduct optical character recognition, are changes in the results of substantive changes in the context or case studied, or a result of the different software used to process the data?

In the context of computational data processing, we define "reproduced processing" as using the same data and the same methods to produce the same data outcomes, despite using different software to conduct the methods. This definition draws conceptually from the definition of reproduction as using the same data and methods and recognizes software as a research tool that should (but may not) produce similar results.

Assessing the reproducibility of transformations to data is not new. There is a large body of research concerning the sensitivity and consistency of text analyses based on choices in preprocessing (Hegazi et al., 2021; Naseem et al., 2021; Tabassum & Patil, 2020). Different choices in text preprocessing have even been

shown to change the performance of downstream models and analyses (Alakrot et al., 2018; Juneja & Das, 2019). For example, Denny and Spirling (2018) show how different choices in pre-processing data for unsupervised learning (such as punctuation removal, lowercasing, stemming, and stopword removal) could produce different latent Dirichlet allocation topic modeling results such that important keywords were only present in the top-20 important terms for a topic for some pre-processed data and not others.

In audio data, the proliferation of methods and tools for analysis poses the same challenges as text. More specifically, there is a need for methods to validate audio analyses. This manuscript proposes a method for validating audio-as-data processing tools, a step in the audio data analysis pipeline.

2.3. Conservative Talk Radio

We apply these strategies to conservative talk radio. Though talk radio has long existed as a form of mediated communication (Armstrong & Rubin, 1989; Avery et al., 1978), conservatives within the United States have leveraged this media format to gain popularity and motivate political action (Matzko, 2020; Young, 2020). However, these tools apply to other audio data, particularly other forms of spoken language (e.g., podcasts, interviews, interpersonal communication, voicemails, and digital videos).

Talk radio combines elements of "shock jock" broadcast entertainment with the promotion of conservative viewpoints, usually hosted by an individual charismatic host. By forming talking points reinforced and amplified within conservative newspapers and on right-wing television, talk radio hosts effectively attack liberal ideas and defend conservatism to their audiences, ultimately driving the rise in outrage as a political media style (Berry & Sobieraj, 2013). Talk radio, as the name suggests, predominantly consists of spoken language by one or several speakers, but it will also include music without words, particularly as brief transitions between sections or as the background of advertising content and introductions. Many talk radio shows, including those studied here, are several hours long and are broadcast daily.

While there have been many popular conservative talk radio hosts, two epitomize the medium: Rush Limbaugh and Alex Jones. Limbaugh, dominant in the 1990s, influenced public opinion (Barker, 1998; Hall & Cappella, 2002; Jamieson et al., 1998; Lee & Cappella, 2001) and conservative rhetoric (Harris et al., 1996). He paved the way for contemporary right-wing media figures, including Glenn Beck and Sean Hannity, many of whom have evolved beyond talk radio into distributing their audio content through podcasting (e.g., Dowling et al., 2022) or YouTube (e.g., Wurst, 2022). Conversely, Jones, known for hosting The Alex Jones Show and Infowars (https://www.infowars.com) since 1999, shaped the conspiratorial wing of modern conservatism in the United States (Beauchamp, 2016), using his platform to propagate conspiracy theories, leading to significant legal and social repercussions (Slater, 2022). His "dangerous demagoguery" (Mercieca, 2019) fosters a unique bond with his audience, affecting trust and media interaction (Madison et al., 2019, 2020), thereby contributing to a fragmented reality among his listeners (Dunne-Howrie, 2019).

We use these two conservative talk show hosts as cases because of their social and political relevance and substantial amount of content: three hours daily. They provide vast and rich corpora of speech-language, making them ideal for comparing and validating audio-as-data tools. Another advantage of talk radio is that it is professionally produced, making it similar to traditional radio, podcasts, and broadcast television content. Audio data produced professionally tends to have greater clarity, as the speakers use professional equipment

and background noise may be minimized. In contrast, both reductive and additive processing may be more difficult in informal recordings with significant background noise. Thus, we expect these results to be most relevant to professionally produced audio media with limited music.

Based on the above literature, we propose the following two research questions:

> RQ1: When processing talk radio audio data into transcripts, to what extent will two different tools for automated transcription reproduce similar data?
>
> RQ2: When processing talk radio audio data to annotate audio features, to what extent will two different tools for audio feature annotation reproduce similar data?

3. Methods

3.1. Data Collection

For our analysis of audio data, we collected data from two right-wing media figures: Alex Jones and Rush Limbaugh. First, the primary data source for our audio recordings of The Alex Jones Show was the official Infowars website's RSS section, specifically titled "Infowars Audio/Video Resource Links." This digital archive, accessible to the public, offers a vast collection of episodes spanning multiple years. In this article, we downloaded all shows from January 1, 2016 to December 31, 2018. We select this time period as Jones received significant and negative attention for his talk shows during this time (in late 2018, Jones was sued by several parents of victims of the Sandy Hook shooting, leading to greater scrutiny of his show; see Williamson, 2022). Each episode within this period adheres to a distinct and consistent URL format, integrating the broadcast date and the day of the week. We employed a custom-built Python tool built on the urllib Python library to ensure a systematic and efficient data retrieval process. This software was designed to: (a) navigate sequentially through the dates from the start of 2016 to the end of 2018, (b) dynamically generate the appropriate URL for each episode based on the date, and (c) initiate the download of the associated audio file, ensuring data integrity and completeness. We sampled 100 episodes from this three-year period to use in the analysis.

Next, we collected audio data for The Rush Limbaugh Show. We used Python code and scraped the audio files hosted on the website (https://www.rushlimbaugh.com) in October 2022. Due to data availability, we could download all Rush Limbaugh shows between January 1, 2020, and June 30, 2021. We randomly sampled 50 shows from 2020 and 50 from 2021 for The Rush Limbaugh Show.

3.2. Reductive: Transforming Audio Data to Text Data

After downloading the audio transcripts, we performed automated speech-to-text using two tools: Google Cloud Platform (GCP) Speech-to-Text and OpenAI's Whisper speech-to-text processing. First, we used Google Speech-to-Text, a leading automatic speech recognition tool (Shakhovska et al., 2019), to generate transcripts of the audio recordings. After experimenting with various models, we settled on Google's "latest_long" with default parameters. To identify different voices within the recordings, we incorporated a diarization configuration, setting the speaker count to range between 2 and 10. The transcription process was automated using a Python script. This approach also significantly reduced the overall processing time.

Compared to GCP Speech-to-Text, OpenAI Whisper is newer. However, it has already been applied to study a variety of communications, including political discourse (Bianchini et al., 2023) and social media audio (e.g., Sihag et al., 2023). Whisper is an "automated speech recognition" tool for multiple languages (Radford et al., 2023). Like other OpenAI tools, Whisper leverages large neural networks—in this case, for conducting automated transcription. The Whisper package in Python contains five models varying in size (and, by extension, speediness and processing requirements). We use the base, English-only model as we anticipate that neither Limbaugh nor Jones would have much non-English in their shows. Open AI's Whisper transcriber is imperfect, but it has achieved significant milestones to make it more applicable to human subject research. We keep these scales at cost by minimizing machine translation for contextualization; however, we acknowledge the process is also subjective.

Owing to cost (while Open AI's Whisper is free, the cost of GCP's Speech-to-Text was about US$50 per transcript, with an overall transcription cost of US$4,106), we only conducted this comparison on Alex Jones for this analysis. In addition to these two specific tools, there are other popular speech-to-text processors that we did not consider due to costs, including Amazon Transcribe (which is known to suffer from long processing times), Microsoft Azure Cognitive Services (which is the costliest for very long recordings, such as talk shows), and IBM Watson Speech to Text (which is more sensitive to more noisy data). While our analysis is exclusively focused on the spoken English language, it is worth noting that of these tools, GCP Speech-to-Text can support a greater variety of languages.

3.2.1. Tool Comparison

To compare the speech-to-text results of GCP's Speech-to-Text with Open AI's Whisper, we compared the transcripts using word error rates (WERs; Klakow & Peters, 2002). WER is a metric that quantifies the difference between the two documents—in this case, the two transcriptions. It measures the percentage of words in the reference transcription incorrectly recognized or omitted in the system's output. Importantly, this is not a measure of understanding (Wang et al., 2003) but a measure of similarity; that is, to what extent are the data similar when the procedure is reproduced across these two tools? We calculated a WER for each episode using the {wersim} package (Proksch et al., 2019). A lower WER is typically better, whereas a higher WER suggests differences in the two transcripts. Previous work has considered WERs of 0.70 to be different (e.g., Jeanrenaud et al., 1995), and more recent studies have considered WERs of 0.40 to be "high error rates" (Morchid et al., 2016, p. 76). This analysis aims to compare the results of Whisper and GCP Speech-to-Text.

An important caveat here involves the data processing for Whisper: WhisperAI has a maximum size of 25 MB. As most of Alex Jones's shows are long (most MP3 recordings were about 55 MB), it was necessary to split the data into three files before proceeding with the analysis. Once these were transcribed, we then combined the three files.

3.3. Additive: Annotating Audio Features

We additionally compare and validate tools that enhance and annotate audio data with information about pitch and intensity, two important metrics in signal processing and auditory analysis (Samrose & Hoque, 2021). To make this comparison, we focus on two packages: parselmouth (Jadoul et al., 2018) and librosa (McFee et al., 2015).

Parselmouth is a Python interface for praat, one of phonology's most popular computer software tools (Kinoshita, 2015; Loakes & Keith, 2013). Praat—and by extension, parselmouth—detects various audio signals, including pitch, tonal intensity, loudness, formants, pacing, and timbre. This tool makes studying accents, prosody (the pacing of someone's speech), and other spoken language phenomena especially useful. For example, Lukito, Gursky, et al. (2023) used praat to study in-person rhetoric at a far-right QAnon event.

Librosa is a tool used for auditory signal processing. Whereas praat and parselmouth are common in linguistics, librosa appears to be more popular in computer engineering and signal processing research to detect sociolinguistic features such as emotion (Babu et al., 2021) and sarcasm detectors (e.g., Tomar et al., 2023).

While some other tools can conduct audio analysis, these two were selected because they specialize in analysis. Other relevant pages include Pysptk, which is also used for speech synthesizing (the artificial production of spoken language) and TorchAudio, which is built on top of PyTorch and, while useful, is more geared towards preparing audio data for machine learning rather than extracting pitch and intensity accurately.

We focus on intensity and pitch detection for this analysis because both are available in parselmouth and librosa, and because of their popularity in signal processing and auditory research. Intensity is most commonly measured using the root mean square amplitude of a sound wave (parselmouth's function is "intensity" and librosa's is "rms," which stands for root mean square). The larger the root mean square, the more intense (or louder) the sound.

Compared to intensity, pitch detection is more challenging. In signal processing and computational phonology, there is no definitive algorithm or way to calculate pitch (Verteletskaya et al., 2009). Correspondingly, these two packages take two different approaches: In parselmouth, pitch is derived from the approximate lowest frequency of the waveform (also known as the fundamental frequency), whereas in librosa, pitch is derived from the melspectogram, an approximation of the waveform's amplitude and frequency over time. In both cases, a higher score constitutes a higher pitch.

3.3.1. Tool Comparison

To compare parselmouth and librosa, we use both to detect the same auditory features (i.e., intensity and pitch) and compare the features extracted with each. For this analysis, we ran the Alex Jones and Rush Limbaugh samples through both. Our goal is to compare the results of parselmouth and librosa for the same transcripts. Notably, both parselmouth and librosa's outputs used different levels of time aggregation, resulting in different numbers of time points. More specifically, parselmouth's analysis was often more granular than librosa's, resulting in typically two to three more time points. Additionally, to make matters more complicated, parselmouth relied on different levels of aggregation for both its mel-spectrometer (for pitch detection) and its intensity detection.

Because these strategies resulted in different temporal aggregation levels, we apply dynamic time warping (DTW) to compare the librosa and parselmouth processing results. DTW is a common approach to compare two-time series that do not perfectly synchronize or vary in speed (Berndt & Clifford, 1994). DTW models leverage non-linear mapping to identify a minimized distance between two or more time series. In DTW,

distance is measured on a scale of 0 to 1, with 0 being *perfect alignment* (no distance, or difference, between the time series) and 1 being *no alignment* (i.e., substantive differences between the time series).

To prepare the data for DTW, we first normalized the time series data (Shao, 2015) to ensure appropriate comparisons of scale between the two-time series. We then conduct four DTW comparisons using the R package {IncDTW} (Leodolter et al., 2021), which provides a vector-based approach to DTW and greater computational efficiency. This step is important for longer-form content, such as podcasts, which create long and highly granular time series data (compared to short online videos and clips). The four DTW comparisons are: (a) a comparison of intensity in Alex Jones's Infowars recordings, (b) a comparison of pitch in Alex Jones's Infowars recordings, (c) a comparison of intensity in Rush Limbaugh's recordings, and (d) a comparison of pitch in Rush Limbaugh's recordings. Below, we present the results for the intensity comparisons and then the pitch comparisons, which are measured as the average distance between each of the point pairs in the two-time series.

4. Results

Our collection consists of 100-episode samples from two sources: Alex Jones's Infowars from 2016 to 2018 and The Rush Limbaugh Show from 2020 to 2021. Both shows are roughly three hours long, totaling 600 hours (about 300 for Jones and Limbaugh each). We begin with comparing the speech-to-text tools for Alex Jones (GCP Speech-to-Text and OpenAI's Whisper), using WER to assess differences in the transcripts. We then compare the computational phonology tools (librosa and parselmouth) using DTW to understand whether these tools produced different auditory features.

4.1. Transcription Results

To address RQ1, we compare the results of GCP's Speech-to-Text with OpenAI's Whisper using WERs. For Alex Jones, the average WER across the 100 videos was 0.76 (SD = 0.21), suggesting a substantial difference between the transcripts produced by GCP Speech-to-Text and Whisper (to compare, the WER for two different episodes was 0.95). These results are presented in Table 1.

A qualitative assessment of the recordings suggests several reasons for this. First, the data splitting ultimately impacted the transcription process despite no content removal. Sometimes, Whisper would not transcribe the first few words of each respective section, resulting in some discrepancies.

A second reason may be related to Whisper's default identification of syntax structure (in particular, punctuations and proper nouns). Whereas automated punctuation requires a secondary model (that was not used in this analysis), Whisper's automated transcription provided fairly accurate punctuations and identification of proper nouns, making it more likely to transcribe names of public figures and groups correctly. We illustrate with an example below (Table 2), from an episode in 2016.

Table 1. WER metrics for Alex Jones data.

Metric	*M*	*SD*	Min	Max
WER	0.762	0.214	0.54	0.89

Table 2. Comparison of GCP Speech-to-Text and OpenAI Whisper.

GCP Speech-to-Text	OpenAI Whisper
Leon McAdoo this put together a powerful six-minute report today that was going to be airing on the Nightly News tonight it's still will we're going to be premiering that here but also be on with more tonight than she's going to Flint Michigan and this evening or tomorrow morning soo	*Leanne* McAdoo has put together a powerful six minute report today that was going to be airing on the nightly news tonight. *And* it still will. We're going to be premiering that here, but *we'll* also be on with more tonight. And she's going to Flint, Michigan this evening or tomorrow morning.

It is important to note that while Whisper appeared to have more accurate punctuations and proper noun identification, both were relatively inconsistent when it came to individual words that are also in portmanteaus (e.g., "takeaway" the noun and "take away" the verb and adverb).

A third reason is that both tools transcribed filler words (e.g., "you know," "ok"), interjectory words (e.g., "wow," "yuck"), and repeated words (e.g., "let, let me start..") differently and inconsistently, which was particularly problematic for sections involving two speakers. GCP was more likely to include repeated and interjectory words, including those spoken softly and in the background. However, this was also inconsistent: There were places where Whisper had identified repeated words, but GCP had not.

4.2. Computational Phonology Results

To address RQ2, we now turn to our analysis of the auditory feature annotations from parselmouth and librosa. Given the differences in how these two tools process the audio data, and in alignment with our methods section, we use DTW to compare the output of these two tools.

Generally, we find that librosa and parselmouth have similar detections of intensity. For the Alex Jones data, the average distance between the two-time series, across our 100-podcast sample, was 0.30 (SD = 0.04), suggesting that librosa and parselmouth had similar identifications of intensity, though these were not identical. Similarly, the average distance between the librosa-derived and parselmouth-derived intensity time series for Rush Limbaugh was 0.27 (SD = 0.014), suggesting that these findings are consistent across podcasts. By comparison, the DTW for intensity of two Alex Jones shows (distance = 0.68) and two Rush Limbaugh shows (distance = 0.74) was very high.

The difference between librosa and parselmouth for pitch is similar. In the case of Alex Jones's data, we find that the average distance between librosa and parselmouth's normalized measure of pitch is 0.30 (SD = 0.04). For Rush Limbaugh, the average distance is 0.29 (SD = 0.016). For comparison, we calculated DTW for intensity of two different Alex Jones shows (distance = 0.77) and two different Rush Limbaugh shows (distance = 0.82). These results are presented in Table 3.

As expected for both pitch and intensity, we find that parselmouth's output is more granular than librosa. Figure 1 illustrates this with a 10-second normalized sample comparison of how parselmouth (red) and librosa (blue) operationalized auditory intensity (from the January 19, 2016 Infowars recording). In this figure, "index," the horizontal axis, refers to the length of the time series (parselmouth's output has more time points and will therefore have a longer index). The y-axis refers to the intensity value as measured by parselmouth or librosa.

Table 3. DTW metrics for Alex Jones and Rush Limbaugh data.

Radio host	Feature	*M*	*SD*	Min	Max
Jones	Pitch	0.306	0.04	0.251	0.322
Jones	Intensity	0.308	0.043	0.262	0.347
Limbaugh	Pitch	0.295	0.016	0.279	0.315
Limbaugh	Intensity	0.271	0.014	0.255	0.289

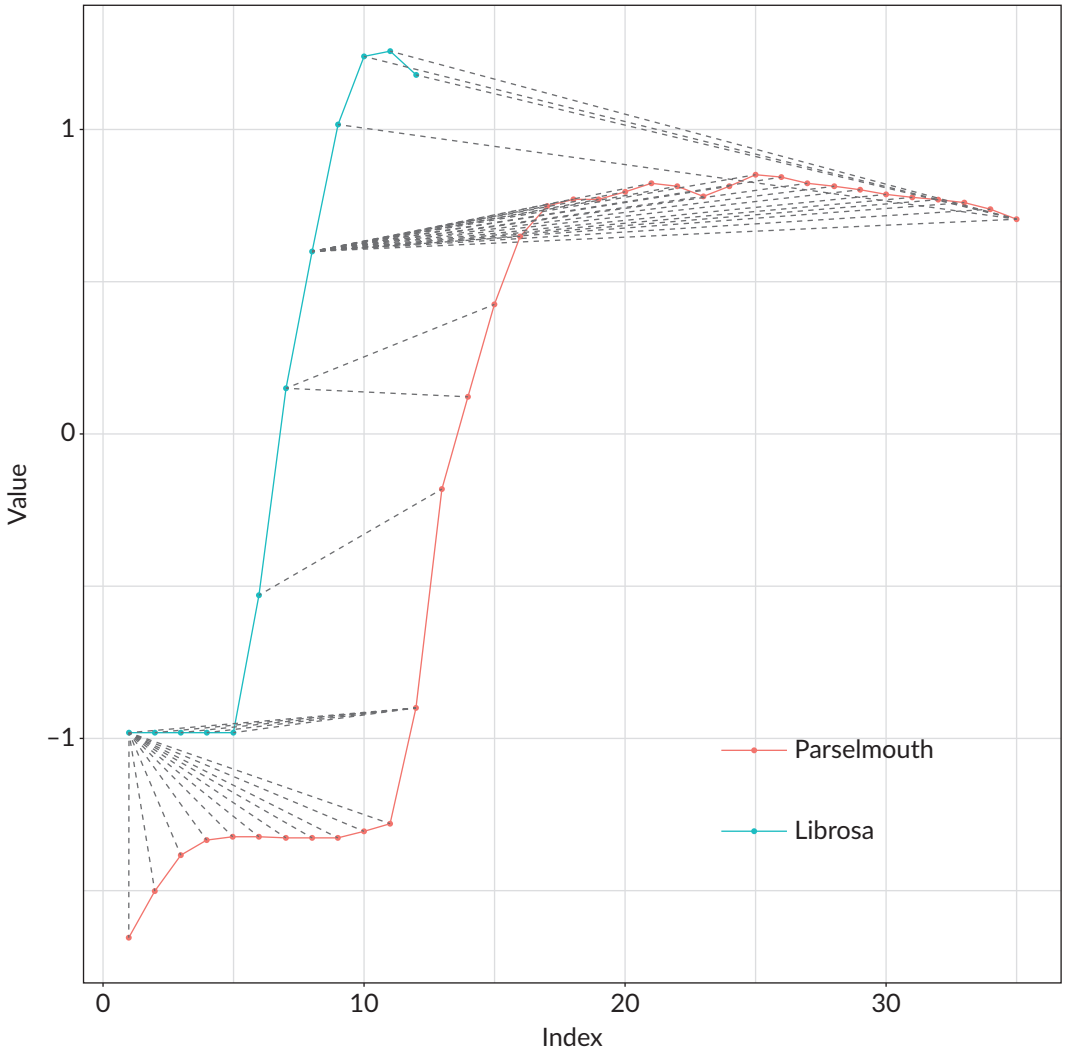

Figure 1. Ten-second time series of intensity as measured by parselmouth and librosa.

The dotted lines between the parselmouth and librosa lines refer to the optimal alignment. In the case of this data, one librosa point may be illustrated by many parselmouth points (the first librosa point on the bottom left has 11 parselmouth points, for example).

While this granularity may be useful when researchers are identifying a millisecond-level analysis (such as with fMRI or physiological phenomena; see Jahn et al., 2022; Lang et al., 2009), these differences may not be as substantive at higher levels of aggregation, such as at the daily or monthly level.

5. Discussion

Our comparison of the speech-to-text and computational phonology tools suggests several differences. Regarding automated transcription, the WER indicates that GCP Speech-to-Text and OpenAI's Whisper may differ. Notably, GCP appears to provide more direct transcriptions, including stutters and fillers, whereas Whisper's transcriptions may have these auditory features removed. Based on these findings, we suggest that researchers select a transcription tool that aligns best with their question or the type of spoken language they seek to study. For example, researchers studying anxiety and pausing in interpersonal conversation may want to identify filler words. However, studies of scripted broadcasts, which already lack many of these language features, may be better suited for processing by OpenAI's Whisper. We also encourage researchers to consider the cost difference between these two tools and whether the increased price for GCP justifies its use.

The comparison of librosa and parselmouth indicates that these two tools were more similar, at least regarding pitch and intensity. However, we also found that parselmouth provided more granular data. This difference has its benefits and disadvantages: Millisecond-level analyses may be necessary for some types of data (e.g., fMRI). However, the more granular output also results in larger and more computationally intensive datasets.

These results highlight the importance of reproduced processing for audio-as-data. While the results produced by the two automated transcription tools (the reductive processing approach) and by the audio feature extraction tools (the additive processing approach) were similar, our study also found some key differences that may motivate researchers to use one tool over the other. This is an essential consideration for reproduction and replication studies as results may not reproduce, not because of the data or context, but because of a difference in processing. Such findings also highlight the importance of methodological transparency and being specific about what packages or programs a researcher used to process their audio data.

In conducting this work, we contribute to the growing literature on data processing (e.g., Denny & Spirling, 2018; Tabassum & Patil, 2020)—and, specifically, its reproducibility—and expand it in the context of audio-as-data. By showcasing methods for reductive and additive processing, and by comparing several tools, we hope this work motivates other media and communication scholars to study audio data more.

Based on these findings, we make several important recommendations for improving the replicability and reproduction of research using audio-as-data. First, reproduction studies should use the same processing approaches, if possible. It is not only a matter of using the same methods, but being explicit about the specific tool, software, programming language, or package/library used to conduct the analysis. Because of the myriad of ways researchers can prepare their data—particularly in computational research and content analysis methods—the ability to replicate a finding is contingent on methodological clarity.

This result leads to our second recommendation: Researchers must also be transparent in their processing approach, including providing information about the tools they used, the version of that tool (if relevant), and the specific steps they conducted in data preparation. This recommendation aligns with open science practices (Dienlin et al., 2020), particularly regarding making research materials more open.

Finally, researchers should validate their results with reproducible processing (i.e., using different processing software to achieve the same task). This procedure ensures the robustness of one's results. Future studies can build on this work by analyzing data processing tools across different types of media (e.g., podcasts, social media content, broadcasts of speeches). For example, one area that would benefit from a greater assessment is the consideration of audio-as-data with music features. While some parts of talk radio contain music (e.g., advertisements), most talk radio audio is spoken language; as such, a limitation of our study is its focus on spoken language. Future studies should seek to reproduce these results with media content that contains music audio, including recordings of songs and movies. Another limitation of this study is our focus on English-language talk radio. As GCP Speech-to-Text claims to support 125 language variations, and OpenAI Whisper claims to support 99, future work can and should consider how these tools may expand non-English audio-as-data studies.

Acknowledgments

We would like to express our gratitude to the reviewers in *Media and Communication* who provided feedback on earlier drafts of the manuscript, as well as the support of the Media and Democracy Data Cooperative members. Ross Dalhke is supported by graduate fellowship awards from Knight-Hennessy Scholars and Stanford Data Science Scholars at Stanford University.

Funding

Funding for this work was provided by the John S. and James L. Knight Foundation.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Alakrot, A., Murray, L., & Nikolov, N. S. (2018). Towards accurate detection of offensive language in online communication in Arabic. *Procedia Computer Science*, *142*, 315–320.

Armstrong, C. B., & Rubin, A. (1989). Talk radio as interpersonal communication. *Journal of Communication*, *39*(2), 84–94.

Avery, R., Ellis, D., & Glover, T. (1978). Patterns of communication on talk radio. *Journal of Broadcasting & Electronic Media*, *22*(1), 5–17.

Babu, P., Nagaraju, V., & Vallabhuni, R. (2021). Speech emotion recognition system with librosa. In G. S. Tomar & K. Sudhakar (Eds.), *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 421–424). IEEE.

Barker, D. (1998). Rush to action: Political talk radio and health care (un) reform. *Political Communication*, *15*(1), 83–97.

Beauchamp, Z. (2016, December 7). Alex Jones, Pizzagate booster and America's most famous conspiracy theorist, explained. *Vox*. http://www.vox.com/policy-and-politics/2016/10/28/13424848/alex-jones-infowars-prisonplanet

Benoit, W., & Holbert, R. (2008). Empirical intersections in communication research: Replication, multiple quantitative methods, and bridging the quantitative–qualitative divide. *Journal of Communication*, *58*(4), 615–628.

Berndt, D., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In U. M. Fayyad & R. Uthurusamy (Eds.), *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining* (pp. 359–370). AAAI Press.

Berry, J. M., & Sobieraj, S. (2013). *The outrage industry: Political opinion media and the new incivility*. Oxford University Press.

Bianchini, G., Zanotti, L., & Meléndez, C. (2023). *Using OpenAI models as a new tool for text analysis in political leaders' unstructured discourse*. Unpublished manuscript. https://osf.io/preprints/psyarxiv/kdngb/download

Bokhove, C., & Downey, C. (2018). Automated generation of "good enough" transcripts as a first step to transcription of audio-recorded data. *Methodological Innovations*, *11*(2). https://doi.org/10.1177/2059799118790743

Chen, X., Yang, J., Gan, S., & Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. *PLoS One*, *7*(1), Article e30278.

Christenson, P. G., & Lindlof, T. R. (1983). The role of audio media in the lives of children. *Popular Music & Society*, *9*(3), 25–40.

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189.

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., . . . de Vreese, C. (2020). An agenda for open science in communication. *Journal of Communication*, *71*(1), 1–26. https://doi.org/10.1093/joc/jqz052

Dietrich, B., Hayes, M., & O'Brien, D. Z. (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, *113*(4), 941–962.

Dowling, D., Johnson, P., & Ekdale, B. (2022). Hijacking journalism: Legitimacy and metajournalistic discourse in right-wing podcasts. *Media and Communication*, *10*(3), 17–27.

Dunne-Howrie, J. (2019). Crisis acting in the destroyed room. *Performance Research*, *24*(5), 65–73.

Gold, B., Morgan, N., & Ellis, D. (2011). *Speech and audio signal processing: Processing and perception of speech and music*. Wiley.

Hall, A., & Cappella, J. N. (2002). The impact of political talk radio exposure on attributions about the outcome of the 1996 US presidential election. *Journal of Communication*, *52*(2), 332–350.

Harris, C., Mayer, V., Saulino, C., & Schiller, D. (1996). The class politics of Rush Limbaugh. *The Communication Review*, *1*(4), 545–564. https://doi.org/10.1080/10714429609388278

Hegazi, M. O., Al-Dossari, Y., Al-Yahy, A., Al-Sumari, A., & Hilal, A. (2021). Preprocessing Arabic text on social media. *Heliyon*, *7*(2), Article e06191. https://doi.org/10.1016/j.heliyon.2021.e06191

Howell, E. (2020). Science communication in the context of reproducibility and replicability: How nonscientists navigate scientific uncertainty. *Harvard Data Science Review*, *2*(4). https://doi.org/10.1162/99608f92.f2823096

Indrayani, Asfiati, S., Riky, M. N., & Rajagukguk, J. (2020). Measurement and evaluation of sound intensity at the Medan Railway Station using a sound level meter. *Journal of Physics: Conference Series*, *1428*(1), Article 012063. https://doi.org/10.1088/1742-6596/1428/1/012063

Iosad, P. (2015). "Pitch accent" and prosodic structure in Scottish Gaelic: Reassessing the role of contact. In M. Hilpert, J.-O. Östman, C. Mertzlufft, M. Rießler, & J. Duke (Eds.), *New trends in Nordic and general linguistics* (pp. 28–54). De Gruyter.

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing parselmouth: A Python interface to praat. *Journal of Phonetics*, *71*, 1–15.

Jahn, N. T., Meshi, D., Bente, G., & Schmälzle, R. (2022). Media neuroscience on a shoestring. *Journal of Media Psychology*, *35*(2). https://doi.org/10.1027/1864-1105/a000348

Jamieson, K. H., Cappella, J. N., & Joseph, T. (1998). Limbaugh: The fusion of party leader and partisan mass medium. *Political Communication*, *15*(Suppl. 1), 1–27. https://doi.org/10.1080/10584609.1998.11672652

Jeanrenaud, P., Eide, E., Chaudhari, U., McDonough, J., Ng, K., Siu, M., & Gish, H. (1995, May 9-12). Reducing word error rate on conversational speech from the Switchboard corpus. In *1995 International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 53–56). IEEE.

Juneja, A., & Das, N. (2019). Big data quality framework: Pre-processing data in weather monitoring application. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 559–563). IEEE.

Kinoshita, N. (2015). Learner preference and the learning of Japanese rhythm. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (pp. 49–62). Iowa State University Press.

Klakow, D., & Peters, J. (2002). Testing the correlation of word error rate and perplexity. *Speech Communication*, *38*(1/2), 19–28.

Klofstad, C. A. (2016). Candidate voice pitch influences election outcomes. *Political Psychology*, *37*(5), 725–738.

Klofstad, C. A., Anderson, R., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704.

Knox, D., & Lucas, C. (2021). A dynamic model of speech for the social sciences. *American Political Science Review*, *115*(2), 649–666.

Krumhansl, C., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 739–751.

Lang, A., Potter, R., & Bolls, P. (2009). Where psychophysiology meets the media: Taking the effects out of mass media research. In J. Bryant & M. B. Oliver (Eds.), *Media effects* (pp. 201–222). Routledge.

Larrouy-Maestri, P., Kegel, V., Schlotz, W., van Rijn, P., Menninghaus, W., & Poeppel, D. (2023). Ironic twists of sentence meaning can be signaled by forward move of prosodic stress. *Journal of Experimental Psychology: General*, *152*(9), 2438–2462. https://doi.org/10.1037/xge0001377

Larsen, E., & Aarts, R. (2005). *Audio bandwidth extension: Application of psychoacoustics, signal processing and loudspeaker design*. Wiley.

Lee, G., & Cappella, J. N. (2001). The effects of political talk radio on political attitude formation: Exposure versus knowledge. *Political Communication*, *18*(4), 369–394.

Leodolter, M., Plant, C., & Brändle, N. (2021). IncDTW: An R package for incremental calculation of dynamic time warping. *Journal of Statistical Software*, *99*(9), 1–23. https://doi.org/10.18637/jss.v099.i09

Loakes, D., & Keith, A. (2013). From IPA to praat and beyond. In K. Allan (Ed.), *The Oxford handbook of the history of linguistics* (pp. 123–140). Oxford University Press.

Lucas, C. (2022). *Package "communication": Feature extraction and model estimation for audio of human speech*. CRAN. https://cran.r-project.org/web/packages/communication/communication.pdf

Lukito, J. (2023). *Political language and the computational turn: Political communication report, 2023*. https://doi.org/10.17169/refubium-39046

Lukito, J., Brown, M. A., Dahlke, R., Suk, J., Yang, Y., Zhang, Y., Chen, B., Kim, S. J., & Soorholtz, K. (2023). *The state of digital media data research, 2023*. Media & Democracy Data Cooperative. https://doi.org/10.26153/tsw/46177

Lukito, J., Gursky, J., Foley, J., Yang, Y., Joseff, K., & Borah, P. (2023). "No reason[.] [i]t /should/ happen here": Analyzing Flynn's retroactive doublespeak during a QAnon event. *Political Communication*, *40*(5), 576–595. https://doi.org/10.1080/10584609.2023.2185332

Luz, S., Masoodian, M., Rogers, B., & Deering, C. (2008). Interface design strategies for computer-assisted speech transcription. In N. Bidwell (Ed.), *Proceedings of the 20th Australasian Conference on Computer–Human Interaction: Designing for Habitus and Habitat* (pp. 203–210). Association for Computing Machinery.

Madison, T. P., Covington, E. N., Wright, K., & Gaspard, T. (2019). Credibility and attributes of parasocial relationships with Alex Jones. *Southwestern Mass Communication Journal*, *34*(2). https://doi.org/10.58997/smc.v34i2.45

Madison, T. P., Wright, K., & Gaspard, T. (2020). "My superpower is being honest:" Perceived credibility and parasocial relationships with Alex Jones. *Southwestern Mass Communication Journal*, *36*(1), 50–64.

Matzko, P. (2020). *The radio right: How a band of broadcasters took on the federal government and built the modern conservative movement*. Oxford University Press.

McEwan, B., Carpenter, C., & Westerman, D. (2018). On replication in communication science. *Communication Studies*, *69*(3), 235–241.

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in Python. In K. Huff & J. Bergstra (Eds.), *Proceedings of the 14th Python in Science Conference* (Vol. 8, pp. 18–25). SciPy.

Mercieca, J. (2019). Dangerous demagogues and weaponized communication. *Rhetoric Society Quarterly*, *49*(3), 264–279.

Morchid, M., Dufour, R., & Linarès, G. (2016). Impact of word error rate on theme identification task of highly imperfect human–human conversations. *Computer Speech & Language*, *38*, 68–85.

Naseem, U., Razzak, I., & Eklund, P. (2021). A survey of pre-processing techniques to improve short-text quality: A case study on hate speech detection on Twitter. *Multimedia Tools and Applications*, *80*, 35239–35266.

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. https://nap.nationalacademies.org/catalog/25303/reproducibility-and-replicability-in-science

Nosek, B. A., & Errington, T. (2020). What is replication? *PLoS Biology*, *18*(3), Article e3000691.

Piñeiro-Otero, T., & Pedrero-Esteban, L.-M. (2022). Audio communication in the face of the renaissance of digital audio. *El Profesional de la Información*, *31*(5). https://doi.org/10.3145/epi.2022.sep.07

Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics*, *11*(76). https://doi.org/10.3389/fninf.2017.00076

Proksch, S., Wratil, C., & Wäckerle, J. (2019). Testing the validity of automatic speech recognition for political text analysis. *Political Analysis*, *27*(3), 339–359.

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 206–219.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th International Conference on Machine Learning* (pp. 28492–28518). PMLR.

Samrose, S., & Hoque, E. (2021). Quantifying the intensity of toxicity for discussions and speakers. In A. Leontyev, T. Yamauchi, & M. Razavi (Eds.), *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 1–5). IEEE.

Shah, D. V., Sun, Z., Bucy, E. P., Kim, S. J., Sun, Y., Li, M., & Sethares, W. (2023). Building an ICCN multimodal classifier of aggressive political debate style: Towards a computational understanding of

candidate performance over time. *Communication Methods and Measures*, *18*(1), 30–47. https://doi.org/10.1080/19312458.2023.2227093

Shakhovska, N., Basystiuk, O., & Shakhovska, K. (2019). Development of the speech-to-text chatbot interface based on Google API. In M. Emmerich, V. Lytvyn, I. Yevseyeva, V. Basto-Fernandes, D. Dosyn, & V. Vysotska (Eds.), *MoMLeT&DS—2019: Modern Machine Learning Technologies and Data Science Workshop* (pp. 212–221). CEUR Workshop Proceedings. https://shorturl.at/elwBO

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, *110*(512), 1797–1817.

Sihag, M., Li, Z. S., Dash, A., Arony, N. N., Devathasan, K., Ernst, N., Branzan Albu, A., & Damian, D. (2023). A data-driven approach for finding requirements relevant feedback from TikTok and YouTube. In K. Schneider, F. Dalpiaz, & J. Horkoff (Eds.), *2023 IEEE 31st International Requirements Engineering Conference (RE 2023)* (pp. 111–122). IEEE.

Slater, J. (2022, October 12). Connecticut jury orders Alex Jones to pay nearly $1 billion to Sandy Hook families. *The Texas Tribune*. https://www.texastribune.org/2022/10/12/alex-jones-sandy-hook-shooting

Spinelli, M., & Dann, L. (2019). *Podcasting: The audio media revolution*. Bloomsbury.

Tabassum, A., & Patil, R. (2020). A survey on text pre-processing & feature extraction techniques in natural language processing. *International Research Journal of Engineering and Technology (IRJET)*, *7*(6), 4864–4867.

Tomar, M., Tiwari, A., Saha, T., & Saha, S. (2023). Your tone speaks louder than your face! Modality order infused multi-modal sarcasm detection. In A. El Saddik, T. Mei, & R. Cucchiara (Eds.), *Proceedings of the 31st ACM International Conference on Multimedia* (pp. 3926–3933). Association for Computing Machinery.

Van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Toward open computational communication science: A practical road map for reusable data and code. *International Journal of Communication*, *13*, 3935–3954.

Verteletskaya, E., Sakhnov, K., & Simak, B. (2009). Pitch detection algorithms and voiced/unvoiced classification for noisy speech. In *2009 16th International Conference on Systems, Signals and Image Processing* (pp. 1–5). IEEE.

Wang, Y., Acero, A., & Chelba, C. (2003). Is word error rate a good indicator for spoken language understanding accuracy. In J. Bilmes & W. Byrne (Eds.), *2003 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 577–582). IEEE.

Williamson, E. (2022). *Sandy Hook: An American tragedy and the battle for truth*. Penguin.

Wisniewski, G., Michaud, A., & Guillaume, S. (2020). Phonemic transcription of low-resource languages: To what extent can preprocessing be automated? In D. Beermann, L. Besacier, S. Sakti, & C. Soria (Eds.), *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)* (pp. 306–315). European Language Resources Association.

Wurst, C. (2022). Bread and plots: Conspiracy Theories and the rhetorical style of political influencer communities on YouTube. *Media and Communication*, *10*(4), 213–223.

Young, D. G. (2020). *Irony and outrage: The polarized landscape of rage, fear, and laughter in the United States*. Oxford University Press.

About the Authors

Josephine Lukito (PhD) is an assistant professor at the University of Texas at Austin's School of Journalism and Media and director of the Media and Democracy Data Cooperative. She is also a Senior Faculty Research Affiliate for the Center for Media Engagement.

Jason Greenfield is a research engineer at NYU's Center for Social Media and Politics. He is especially interested in studying instances of online extremism, the radicalization process, and the interplay between humor and hate.

Yunkang Yang (PhD) is an assistant professor at the Department of Communication and Journalism at Texas A&M University where he is also affiliated with the Data Justice Lab. He has a forthcoming book titled *Weapons of Mass Deception: How Right-Wing Media Wage Information Warfare and Undermine American Democracy*.

Ross Dahlke is a PhD candidate at Stanford University in the Department of Communication.

Megan A. Brown is a PhD student at the School of Information at the University of Michigan.

Rebecca Lewis is a Stanford Graduate Fellow and PhD candidate in Communication at Stanford University. She is an expert on disinformation and far-right digital media.

Bin Chen is a doctoral candidate in Journalism and Media at the University of Texas at Austin. His research focuses on political communication, multi-platform research, and computational social science.