



OPEN ACCESS

EDITED BY

Fasih Haider,
University of Edinburgh, United Kingdom

REVIEWED BY

Kubra Cengiz,
Istanbul Technical University, Türkiye
Noor Alhusna Madzlan,
Sultan Idris University of Education, Malaysia

*CORRESPONDENCE

Willem J. M. I. Verbeke
✉ verbeke@ese.eur.nl

RECEIVED 11 May 2024

ACCEPTED 09 September 2024

PUBLISHED 27 September 2024

CITATION

Sharma A and Verbeke WJMI (2024) Influence of gender dimorphism on audience engagement in podcasts: a machine learning analysis of dynamic affective linguistic and paralinguistic features. *Front. Commun.* 9:1431264. doi: 10.3389/fcomm.2024.1431264

COPYRIGHT

© 2024 Sharma and Verbeke. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Influence of gender dimorphism on audience engagement in podcasts: a machine learning analysis of dynamic affective linguistic and paralinguistic features

Amita Sharma¹ and Willem J. M. I. Verbeke^{2*}

¹Department of Operations Research & Quantitative Analysis, Institute of Agri-Business Management, Swami Keshwanand Rajasthan Agricultural University, Bikaner, India, ²Erasmus University, Rotterdam, Netherlands

Effective communication is a crucial objective for business leaders, educators, and politicians alike. Achieving impactful communication involves not only the selection of appropriate words but also proficiency in their delivery. Previous research has frequently examined linguistic, affective linguistic, and paralinguistic features in isolation, thereby overlooking their cumulative impact over time. This study addresses this gap by utilizing a machine learning approach to analyze the dynamic interplay between affective linguistic and paralinguistic features across various episodes of online podcasts. Furthermore, this research incorporates an analysis of gender disparities, acknowledging the dimorphic nature of language and speech across genders. Our findings suggest that accounting for gender when examining the dynamic interactions between affective linguistic and paralinguistic features over time, known as emotional volatility, significantly improves the explanatory power of variations in audience engagement compared to analyses that consider these variables separately.

KEYWORDS

affective linguistic, paralinguistic, podcasts, audience engagement, gender dimorphism

1 Introduction

“And facts are useful. Sometimes. But just as often they lull an audience to sleep” (Berger, 2023 in *Magic Words*).

In recent years, podcasting has experienced a massive surge in popularity. Today, there are over 400 million podcast listeners worldwide and over 2 million independent podcasts (Backlinko, 2024). Insider Intelligence reports that podcasting will be a \$94.88 billion industry by 2028 (Insider Intelligence, 2024). Not only are the audience numbers growing exponentially, but many communication experts believe podcast presenters are becoming trusted sources of information, which translates into more audience engagement (Brinson and Lemon, 2023).

Given the growing importance of this emerging medium, understanding what drives podcast engagement (which, in this study, is the number of likes and comments from the audience) is what all podcast presenters are eager to know, as it allows them to better craft the spoken messages they seek to get across with their audience. Substantial evidence exists on the individual effects of linguistic features (the words or language used) and paralinguistic features (prosody or acoustic properties of speech, such as pitch and loudness).

Research on verbal communication messages has used a variety of text analysis tools to study relevant communication outcomes such as word of mouth (Berger et al., 2022), leadership emergence (Truninger et al., 2021), speeches of charismatic business leaders (Niebuhr et al., 2016) and persuasion (Rizzo and Berger, 2023). Some text analysis tools focus on a broad spectrum of linguistic features such as Linguistics Inquiry and Word Count (LIWC; Tausczik and Pennebaker, 2010) or vocabulary builders (Wordify; Hovy et al., 2021). Others focus specifically on affective language features, such as the Evaluative Lexicon (Rocklage et al., 2018), Sentiment Classifiers (Pang et al., 2002), or the National Research Council (NRC) Emotion Lexicon (Mohammad and Turney, 2013).

In this study, we focus on affective linguistic features because, in podcasts, these features are known to significantly impact the effectiveness of the content (Berger et al., 2021, p. 236). A substantial portion of people's reactions to story content, specifically podcasts, are driven by their emotional responses to the sentiments expressed in the story, which in turn drives the popularity of podcasts (Rocklage et al., 2018). Therefore, in the following sections, we use the term "affective linguistic features" and employ the NRC Emotion Lexicon for its operationalization (Mohammad and Turney, 2013).

Furthermore, recent academic research has also examined the impact of paralinguistic or non-lexical features of communication such as paralangage classifiers (PARA; Luangrath et al., 2023), PRAAT (Gangamohan et al., 2016), VoiceVibes (Truninger et al., 2021), voice numerosity (Chang et al., 2023), and vocal tones (Wang et al., 2021). This feature involves, among other things, the modulation of vocal elements such as tone, pitch, rhythm, volume, and articulation rate to express varied emotions and, in doing so, highlight specific key points of the podcast. It should be noted that many authors employ slightly different terms when discussing specific paralinguistic features related to speaking rate, such as speech tempo, articulation rate, or speaking rate. In this study, we use the term "articulation rate" and define it as the speed at which people speak (Quené, 2008; Cascio Rizzo et al., 2023).

It is unclear about how these affective linguistic and paralinguistic features can interact collectively to influence the impact of online verbal communication. This aspect is particularly relevant in the context of podcasts, the primary communication channel investigated in this study.

1.1 Purpose of the study

The current study builds upon existing research into automated text analysis by focusing on both the affective linguistic and paralinguistic features of communication. We explore the interaction effects of these features on communication messages, specifically within the dynamic context of emphatic accents and prosodic episodes or contours. Emphatic accents refer to the production of speech sounds that emphasize certain words or phrases (Novák-Tót et al., 2017), while prosodic episodes or contours describe broader patterns and variations

in speech over time that segment speech to convey specific meanings and emotions (Bänziger and Scherer, 2005). In this context, we refer to these paralinguistic terms collectively as "prosodic contours." By leveraging both these affective linguistic and paralinguistic features, we propose that podcast presenters can significantly enhance their impact, allowing presenters to convey their intentions and key messages more compellingly. Moreover, incorporating affective speech features into specific fragments of a speech not only prevents monotony but also highlights certain parts as more or less prominent or surprising. This adds emotional volatility to the message and thereby enhances audience engagement (Berger, 2023; Berger et al., 2021). The careful segmentation of speeches, a characteristic of charismatic leaders' speeches (Niebuhr et al., 2016), exemplifies this approach.

Furthermore, as the vocal characteristics of individuals convey pertinent details, notably pertaining to gender (Lausen and Schacht, 2018; Ko et al., 2006), and are recognized to elicit variances in gender-specific persuasion effects (Clarke and Healey, 2022), we are examining the gender-dimorphic attributes inherent in language and speech. This study thus explores how the gender of the podcast presenter influences the effects of affective linguistic and paralinguistic features, as well as their dynamic interaction on audience engagement (Schirmer and Kotz, 2006; Novák-Tót et al., 2017).

Our study provides several contributions. First, we examine the influence of affective linguistic and paralinguistic features on verbal communication messages. We demonstrate how affective linguistic features, such as expressions of sadness, interact with paralinguistic features, such as pitch, to influence audience engagement (Schwartz and Pell, 2012). For instance, expressions of sadness delivered with varying pitches may convey different meanings and potentially generate more audience engagement.

Second, during podcasts, we explore the development of dynamic affective linguistic and paralinguistic markers. While existing text analysis tools have focused primarily on the average presence of (affective) linguistic and paralinguistic features in a speech or presentation, we investigate how episodic variations of these features across segments in a speech (e.g., at the beginning, end, or during periods of contrast) impact communication. Our approach, by examining the dynamic interaction between affective linguistic and paralinguistic features over time, provides higher granularity in data analysis, offering insights into how podcast presenters can more effectively craft their narratives so as to add emotional volatility and increase audience engagement (Berger et al., 2021).

Third, we address potential presenter gender effects or gender dimorphism. Previous research has suggested that speaker gender may interact with paralinguistic features to affect communication quality. However, conclusive empirical evidence for this interaction has been lacking (McSweeney et al., 2022; Van Zant and Berger, 2020; Lausen and Schacht, 2018). Our study addresses this gap by demonstrating potential differences in what affective linguistic features and paralinguistic features and their interactions drive audience engagement among male and female podcast presenters.

We also consider the potential impact of podcast length on audience engagement, as previous research suggests that the length of a podcast can influence audience responses (Jeon et al., 2019; Cosimini et al., 2017; Chan-Olmsted and Wang, 2022). Given that audiences often multitask while listening to podcasts (Perks et al., 2019), the length of a podcast must be considered as it may affect the listener's willingness to engage with longer content.

In the sections that follow, we first present a review of existing text analysis research and how affective linguistic and paralinguistic features shape the impact of communication messages. In the next section, we discuss major gaps in existing research on communication message effectiveness and outline how we intend to address them.

1.2 Theory

1.2.1 Verbal communication messages

In verbal communication, messages are meticulously constructed to convey specific information from a sender to a receiver. This may constitute either a tangible audience (e.g., in face-to-face interactions) or an abstract audience (e.g., in podcasts). The sender strategically formulates and crafts these messages through speech in order to achieve desired outcomes or objectives, as elucidated by [Grice \(1957\)](#) and further elucidated by [Pickering and Garrod \(2004\)](#). These objectives may encompass persuasion ([Van Zant and Berger, 2020](#), p. 661), alteration of attitude ([Petty and Cacioppo, 2012](#)), or the motivation to undertake preventive health measures ([O’Keefe and Jensen, 2007](#)). This intentional shaping of verbal messages, extending beyond the mere accumulation of facts, entails the utilization of both affective linguistic features and paralinguistic features to more effectively fulfill the communicator’s aims and intentions ([Rocklage et al., 2018](#)).

This study emphasizes audience engagement, which can be broadly conceptualized as a customer’s motivationally driven voluntary investment in cognitive, emotional, and behavioral resources in response to a communication message. This definition is adapted from [Hollebeek \(2019\)](#) and [Perks et al. \(2019\)](#). In this context, engagement is quantitatively measured by the number of likes and comments generated by a podcast from its audience.

It is noteworthy that this project employs a machine learning approach, utilizing a substantial dataset that includes a wide range of both affective linguistic and paralinguistic features or markers. Consequently, we are inclined to make general conjectures rather than concrete hypotheses. Machine learning is particularly adept at exploring and uncovering novel insights through extensive datasets. Therefore, in this study, we discuss the development of an audience engagement prediction model based on machine learning, which accounts for a significant variance in podcast audience engagement.

1.2.2 Affective linguistic and paralinguistic features

Although previous research has begun to investigate the impact of both overall and affective linguistic features (e.g., words, phrases, or styles of language; [Berger et al., 2022](#); [Tausczik and Pennebaker, 2010](#)) and paralinguistic features (e.g., articulation rate or pitch; [Gangamohan et al., 2016](#); [Truninger et al., 2021](#)), these aspects have mostly been studied separately. There has been less attention given to how they mutually shape communication impact.

In speech, particularly when conveying verbal messages, avoiding monotony and accentuating specific implications beyond words is crucial. Podcast presenters, therefore, infuse their messages with prosodic contours and episodes to enhance communication effects ([Bänziger and Scherer, 2005](#)). Prosodic contours and episodes in communication represent the dynamic expression and variation of emotions or emphasis throughout a message segment. This involves, among other things, modulating vocal elements such as tone, pitch,

rhythm, volume, and articulation rate (or tempo) to express varied emotions and emphasize specific key points. In this study, when referring to using affective linguistic and paralinguistic features over the total podcast length, we refer to “average” features. And, when referring to features or markers across specific segments in the podcast narrative, we refer to “dynamic” or “granular” linguistic and paralinguistic features that accentuate specific sequences. Hence, we propose the following conjecture:

Conjecture 1: Combining dynamic, granular affective linguistic, and paralinguistic features, as well as their interaction across various podcast episodes, enhances the variance explained in audience engagement prediction models compared to using average affective linguistic and paralinguistic features separately.

1.2.3 Gender dimorphism

Gender dimorphism, as it relates to speech, is a fundamental biological process where various biological characteristics, including the vocal folds in the larynx, as well as changes in heart rate, blood flow, and muscle tension that are influenced by arousal—impact speech production ([Schirmer and Adolphs, 2017](#); [Schirmer and Kotz, 2006](#)). These paralinguistic features during speech are referred to as “leakier,” thereby revealing indexical attributes of the speaker and thus potentially impacting listeners’ interpretation of the spoken message ([Pernet and Belin, 2012](#); [van Berkum et al., 2008](#)). However, it is important to note that interpretation bias dynamics are complex; e.g., gender stereotyping and expectations by a listener are culturally dependent and might negatively or positively bias what and how something is being said by male versus female speakers ([Clarke and Healey, 2022](#); [Lausen and Schacht, 2018](#)).

Females typically exhibit a higher pitch than males, also known as fundamental frequency ([Leung et al., 2018](#)), a phenomenon that can be attributed to physiological differences in the vocal cords ([Pernet and Belin, 2012](#); [Daly and Warren, 2001](#)). This higher pitch, however, also enables females to modulate their pitch more effectively during speech, which might enhance the expressiveness, passion, and emotional nuances of their spoken language ([Pisanski et al., 2018](#)). Additionally, females generally speak more rapidly than males and employ shorter pauses, a pattern that correlates with the higher pitch of the female voice. Moreover, women tend to use more varied intonation patterns, making their speech more dynamic and often employing rising intonation, e.g., beginning versus the end of a sentence. However, this might give the audience the impression that the speaker is nervous or has a lack of confidence ([Henton, 1995](#)). This modulation, however, might evoke a greater engagement by the listener and signal speakers’ openness to feedback.

Conversely, males typically exhibit a lower pitch and frequently employ longer pauses, which may convey authority ([Aung and Puts, 2020](#); [Tigue et al., 2012](#)). This insight could then motivate females to lower their pitch, resulting in sex-atypical or artificial speech ([Anderson et al., 2014](#)). It is imperative to note that these observations reflect general or average characteristics of speech across diverse speech delivery situations or contexts.

Nonetheless, these gender-dimorphic traits can potentially bias listeners in particular contexts. For instance, [Clarke and Healey \(2022\)](#) highlight several instances wherein female entrepreneurs may encounter increased challenges in persuading investors owing to their naturally higher voice pitch (also see [Balachandra et al., 2019](#); [Niebuhr et al., 2019](#)). Conversely, other researchers, such as [Li et al. \(2022\)](#),

have demonstrated in crowdfunding campaigns that various characteristics, besides pitch, such as using friendly words, create differences between male and female speakers and influence persuasiveness for both genders. The focus of this study, however, is on podcast presenters, who are known for carefully crafting their messages. [Berry and Brown \(2019\)](#) describe this as being “in actor role,” which involves making certain segments of the podcast more salient or crafting the suspense experience within the podcast narrative.

While there is a definitive gender difference in normal speech, it remains unclear whether male or female presenters mark specific affective words or parts of the podcast with distinctive paralinguistic features, such as prosodic contours, during specific segments of a speech. Given the disparities in speech modulation abilities, significant differences are anticipated in the utilization of distinct dynamic affective linguistic and paralinguistic features and their impact on audience engagement.

Conjecture 2: Modeling the audience engagement prediction with affective linguistic and paralinguistic features will separately predict greater variances in audience engagement for males and females.

2 Data and methods

2.1 Data

Data analysis of the audio files is extremely complex and involves higher computing power. We used a cloud server to process the extraction of audio files from the YouTube channel, TEDx Shorts. We selected the TEDx Shorts YouTube channel because of a certain level of uniformity within the dataset. Every video on the YouTube channel has certain attributes, such as thumbnails, the title of the video, the description of the video, the time span of the video (right lower corner of thumbnail), the number of subscribers, the number of views, and the number of likes and comments (see [Table 1](#) for a description of the different audio files).

A thumbnail is a small, low-resolution version of a larger image or video that is displayed on each video on a YouTube channel. This is the first property of a YouTube video that ignites the viewer to watch the full video. YouTube thumbnails have several important functions, such as attracting viewers, summarizing content, branding, increasing click-through rate, engagement, and accessibility. Hence, thumbnails constitute a crucial aspect of YouTube video promotion and can significantly impact a video's success in terms of views, engagement, and audience retention.

In our videos from the TEDx Shorts Podcast YouTube Channel, the thumbnails were uniformly designed; therefore, we can assume that the impact of thumbnails on attractive viewership or click-through rate is negligible. Our study solely focused on video content to understand how affective linguistic, paralinguistic, dynamism, and interaction among the various features extracted from video content make an impact on the engagement of the audience.

2.2 Data extraction

This study considers a sample of 188 TEDx Shorts podcasts downloaded from [Youtube.com](#). A typical TEDx Shorts podcast lasts for an average of 6.18 min and addresses various educational topics in

an entertaining way, hence the name edutainment ([Aksakal, 2015](#); [Oslawski-Lopez and Kordsmeier, 2021](#)). Our sample was divided into 112 female speakers and 76 male speakers. Between 1 Dec 2021 and 1 Dec 2023, the podcasts have an average of 13,245 views and 266 engagements, which is the sum of likes and comments.

The downloading of the 188 recordings of TEDx Shorts podcasts was performed using a Python program. A JSON file was extracted for each podcast. Each JSON file was converted to a comma-separated value (CSV) file. Each CSV file contained three columns, namely, word, start, and end. The words column contained the words spoken in the podcast. The start column stores the time stamp in seconds when the corresponding word was initiated by the speaker, whereas the end column stores the time stamp in seconds when the corresponding word was stopped speaking by the speaker.

We collected two types of data for this research: a CSV file and a Waveform Audio File Format (WAV) for each podcast of the TEDx Shorts Podcast YouTube Channel. Furthermore, we removed the first 45 s of the CSV and WAV files because they contained the introduction signature of the TEDx Shorts and a few lines on the introduction of the speaker.

2.3 Data preparation

The CSV and WAV files of 188 podcasts have different time spans. We created 152 independent variables and three dependent variables (viewCount, likeCount, and commentCount) from the CSV and WAV files of each podcast. We removed the first column from the dataset that contained the podcast titles. Subsequently, the first five columns depicted the number of views (named viewCount), number of likes (named likeCount), number of comments (named commentCount), and the female (representing the gender of the speaker) information of each speaker, which was designated as 1 for female speakers and 0 for male speakers, along with the podcast duration measured in seconds. The viewCount, Gender, and Podcast Length were used as control variables for data analysis and model building. The likeCount and the commentCount were used as outcome variables.

Engagement in a YouTube video refers to the level of interaction and activity that viewers experience with the video content. It includes various metrics that measure how users interact with the video and its associated elements. Common indicators of engagement on YouTube videos include views, watch time, likes and dislikes, comments, shares, subscriptions, and click-through rate. For our research purpose, we defined engagement as an arithmetic sum of likes and comments on the video of the TEDx Shorts Podcast Channel. We used the natural log transformation of the engagement (as shown in [Figure 1](#)).

$$\text{Engagement} = \log(\text{likeCount} + \text{commentCount})$$

The log transformation improved the data normality in comparison to the raw form of Engagement. However, statistically, the log-transformed Engagement (logEng) is not normal, but the Quantile–Quantile plot of logEng shows normality in the data except for the extreme ends. We did not remove the extreme values from the dataset as it could reduce the number of observations.

From each CSV and WAV file of a podcast, 152 variables ([Table 2](#) shows the variables and respective definitions) were extracted and used

TABLE 1 The data source.

| Anonymized title | viewCount (Number of views) | likeCount (Number of likes) | commentCount (Number of comments) | Podcast Length (Length of podcast in seconds) | nWords (Number of words spoken in the podcast) |
|------------------|-----------------------------|-----------------------------|-----------------------------------|---|--|
| Video1 | 4,769 | 32 | 10 | 383 | 1,304 |
| Video2 | 1934 | 18 | 0 | 327.2 | 928 |
| Video3 | 8,564 | 64 | 1 | 364.46 | 1,342 |
| Video4 | 13,042 | 133 | 4 | 388.44 | 1,345 |
| Video5 | 17,097 | 186 | 2 | 486.78 | 1,319 |
| Video6 | 32,455 | 783 | 30 | 341.52 | 965 |
| Video7 | 28,324 | 651 | 1 | 461.86 | 1,165 |
| Video8 | 3,309 | 28 | 2 | 357.16 | 951 |
| Video9 | 5,156 | 58 | 3 | 348 | 1,239 |
| Video10 | 9,819 | 104 | 3 | 330.54 | 908 |
| Video11 | 45,586 | 1,144 | 0 | 295.52 | 991 |
| Video12 | 4,741 | 46 | 0 | 285.04 | 957 |
| Video13 | 2,512 | 29 | 3 | 549.56 | 1716 |
| Video14 | 6,960 | 68 | 3 | 445.14 | 1,473 |
| Video15 | 2,523 | 26 | 2 | 433.44 | 1,268 |
| Video16 | 3,984 | 24 | 1 | 329.68 | 916 |
| Video17 | 8,725 | 36 | 0 | 342.8 | 1,058 |
| Video18 | 1986 | 33 | 4 | 428.36 | 1,273 |
| Video19 | 13,988 | 143 | 5 | 309.66 | 1,014 |
| Video20 | 4,197 | 59 | 2 | 425.12 | 1,495 |
| Video21 | 4,449 | 17 | 3 | 348.28 | 1,210 |
| Video22 | 2,789 | 27 | 2 | 427.86 | 1,398 |
| Video23 | 31,859 | 699 | 1 | 355.76 | 1,206 |
| Video24 | 21,743 | 374 | 6 | 437.34 | 1,458 |
| Video25 | 5,597 | 64 | 11 | 383.52 | 1,083 |
| Video26 | 5,065 | 43 | 1 | 395.68 | 1,502 |
| Video27 | 1957 | 39 | 3 | 426.56 | 1,123 |
| Video28 | 6,023 | 39 | 2 | 351.72 | 1,198 |
| Video29 | 2,715 | 32 | 1 | 549.98 | 1,558 |
| Video30 | 21,984 | 508 | 5 | 337.38 | 1,058 |
| Video31 | 5,709 | 28 | 0 | 483.7 | 1,653 |
| Video32 | 2,978 | 24 | 0 | 332.54 | 938 |
| Video33 | 6,325 | 69 | 2 | 365.26 | 1,114 |
| Video34 | 18,185 | 366 | 7 | 239.96 | 725 |
| Video35 | 2,517 | 32 | 3 | 377.08 | 1,159 |
| Video36 | 34,449 | 778 | 2 | 318.78 | 1,038 |
| Video37 | 4,773 | 50 | 1 | 304.02 | 1,012 |
| Video38 | 3,589 | 36 | 4 | 409.68 | 1,274 |
| Video39 | 1,333 | 10 | 6 | 403.28 | 984 |
| Video40 | 20,685 | 211 | 29 | 404.94 | 1,279 |
| Video41 | 1994 | 27 | 1 | 379.42 | 1,042 |
| Video42 | 9,563 | 73 | 57 | 558.5 | 1,557 |
| Video43 | 4,273 | 47 | 2 | 360.58 | 1,056 |

(Continued)

TABLE 1 (Continued)

| Anonymized title | viewCount (Number of views) | likeCount (Number of likes) | commentCount (Number of comments) | Podcast Length (Length of podcast in seconds) | nWords (Number of words spoken in the podcast) |
|------------------|-----------------------------|-----------------------------|-----------------------------------|---|--|
| Video44 | 3,963 | 52 | 30 | 471.16 | 1,358 |
| Video45 | 19,622 | 337 | 44 | 323.4 | 1,074 |
| Video46 | 38,446 | 834 | 3 | 253.74 | 736 |
| Video47 | 3,641 | 48 | 0 | 289.64 | 1,017 |
| Video48 | 8,613 | 92 | 2 | 283.74 | 779 |
| Video49 | 34,605 | 789 | 25 | 481.64 | 1,601 |
| Video50 | 3,601 | 29 | 2 | 375.62 | 980 |
| Video51 | 6,776 | 58 | 2 | 272.42 | 640 |
| Video52 | 33,228 | 772 | 3 | 494.66 | 1816 |
| Video53 | 5,132 | 64 | 1 | 357.42 | 942 |
| Video54 | 3,393 | 26 | 3 | 335.86 | 852 |
| Video55 | 15,760 | 111 | 7 | 284.18 | 892 |
| Video56 | 3,592 | 54 | 0 | 404.98 | 1,405 |
| Video57 | 40,382 | 1,079 | 26 | 376.72 | 1,485 |
| Video58 | 3,086 | 13 | 25 | 483.08 | 1,256 |
| Video59 | 4,891 | 43 | 0 | 396.58 | 1,141 |
| Video60 | 6,370 | 112 | 1 | 501.58 | 1895 |
| Video61 | 6,247 | 78 | 6 | 491.9 | 1,643 |
| Video62 | 10,384 | 77 | 6 | 390.82 | 1,196 |
| Video63 | 2,228 | 22 | 3 | 444.03 | 1,452 |
| Video64 | 44,086 | 917 | 19 | 283.26 | 735 |
| Video65 | 7,555 | 69 | 3 | 405.92 | 1,071 |
| Video66 | 4,365 | 57 | 8 | 391.66 | 1,428 |
| Video67 | 2,539 | 39 | 38 | 465.94 | 1,370 |
| Video68 | 4,956 | 37 | 1 | 254 | 781 |
| Video69 | 5,570 | 90 | 25 | 429.26 | 1,167 |
| Video70 | 9,228 | 136 | 3 | 519.66 | 1950 |
| Video71 | 3,553 | 31 | 0 | 363.3 | 1,006 |
| Video72 | 29,878 | 635 | 1 | 333.74 | 893 |
| Video73 | 7,785 | 113 | 58 | 362.82 | 1,071 |
| Video74 | 30,987 | 700 | 2 | 480.52 | 1,639 |
| Video75 | 7,443 | 126 | 3 | 436.56 | 1,393 |
| Video76 | 3,666 | 54 | 1 | 310.26 | 959 |
| Video77 | 22,108 | 399 | 30 | 332.74 | 1,004 |
| Video78 | 5,305 | 42 | 1 | 229.58 | 731 |
| Video79 | 13,817 | 88 | 5 | 318.08 | 1,038 |
| Video80 | 4,290 | 39 | 31 | 349.28 | 830 |
| Video81 | 29,814 | 625 | 102 | 395.72 | 1,164 |
| Video82 | 41,139 | 894 | 9 | 504.96 | 1,307 |
| Video83 | 3,626 | 40 | 2 | 431.2 | 1,553 |
| Video84 | 2,515 | 29 | 3 | 414.16 | 1,271 |
| Video85 | 4,096 | 32 | 1 | 357.9 | 969 |
| Video86 | 9,049 | 78 | 0 | 303 | 1,114 |

(Continued)

TABLE 1 (Continued)

| Anonymized title | viewCount (Number of views) | likeCount (Number of likes) | commentCount (Number of comments) | Podcast Length (Length of podcast in seconds) | nWords (Number of words spoken in the podcast) |
|------------------|-----------------------------|-----------------------------|-----------------------------------|---|--|
| Video87 | 12,458 | 134 | 4 | 499.14 | 1,316 |
| Video88 | 4,625 | 35 | 6 | 323.94 | 1,115 |
| Video89 | 6,925 | 56 | 6 | 376.38 | 1,104 |
| Video90 | 11,350 | 59 | 1 | 204.3 | 694 |
| Video91 | 5,836 | 52 | 0 | 310.62 | 995 |
| Video92 | 4,672 | 47 | 44 | 550.94 | 2056 |
| Video93 | 1,468 | 12 | NA | 592.26 | 2024 |
| Video94 | 14,161 | 205 | 4 | 378.78 | 1,104 |
| Video95 | 21,629 | 388 | 51 | 349.48 | 933 |
| Video96 | 9,581 | 112 | 320 | 465.52 | 1749 |
| Video97 | 11,013 | 78 | 2 | 322.86 | 1,051 |
| Video98 | 1,542 | 13 | 0 | 431.2 | 1,304 |
| Video99 | 7,057 | 49 | 18 | 406.4 | 1,335 |
| Video100 | 11,512 | 375 | 39 | 327.6 | 991 |
| Video101 | 4,711 | 60 | 31 | 418.66 | 1,209 |
| Video102 | 9,001 | 116 | 15 | 331.02 | 1,181 |
| Video103 | 5,663 | 74 | 10 | 292.78 | 1,097 |
| Video104 | 5,577 | 110 | 12 | 281.2 | 958 |
| Video105 | 3,733 | 58 | 9 | 530.16 | 1,605 |
| Video106 | 36,107 | 480 | 5 | 289.56 | 1,043 |
| Video107 | 9,349 | 111 | 6 | 230.68 | 738 |
| Video108 | 11,716 | 137 | 7 | 292.52 | 799 |
| Video109 | 37,537 | 1,061 | 50 | 236.22 | 941 |
| Video110 | 54,003 | 1,245 | 6 | 347.16 | 1,031 |
| Video111 | 32,448 | 976 | 66 | 355.84 | 913 |
| Video112 | 4,589 | 83 | 5 | 333.2 | 1,016 |
| Video113 | 43,986 | 1,217 | 5 | 345.04 | 1,233 |
| Video114 | 4,930 | 55 | 3 | 405.18 | 1,345 |
| Video115 | 11,844 | 206 | 9 | 292.34 | 1,109 |
| Video116 | 29,544 | 633 | 62 | 291.96 | 979 |
| Video117 | 6,533 | 92 | 1 | 363.26 | 1,138 |
| Video118 | 3,036 | 30 | 3 | 366.44 | 1,196 |
| Video119 | 5,146 | 95 | 12 | 322.5 | 1,029 |
| Video120 | 4,895 | 54 | 3 | 442.46 | 1,357 |
| Video121 | 6,037 | 66 | 3 | 419.92 | 1,300 |
| Video122 | 21,560 | 487 | 2 | 472.5 | 1,427 |
| Video123 | 9,356 | 87 | 8 | 486.42 | 1,694 |
| Video124 | 11,014 | 193 | 4 | 326 | 1,010 |
| Video125 | 7,997 | 97 | 2 | 391.02 | 1,082 |
| Video126 | 14,052 | 169 | 9 | 342.02 | 1,055 |
| Video127 | 5,624 | 73 | 0 | 404.54 | 1,290 |
| Video128 | 3,895 | 63 | 8 | 414.3 | 1,366 |
| Video129 | 8,167 | 96 | 47 | 301.96 | 968 |
| Video130 | 5,337 | 76 | 85 | 297.08 | 720 |

(Continued)

TABLE 1 (Continued)

| Anonymized title | viewCount (Number of views) | likeCount (Number of likes) | commentCount (Number of comments) | Podcast Length (Length of podcast in seconds) | nWords (Number of words spoken in the podcast) |
|------------------|-----------------------------|-----------------------------|-----------------------------------|---|--|
| Video131 | 8,490 | 125 | 168 | 333.8 | 1,084 |
| Video132 | 5,926 | 87 | 43 | 337.74 | 1,072 |
| Video133 | 7,187 | 68 | 13 | 396.7 | 1,590 |
| Video134 | 12,643 | 162 | 11 | 327.22 | 1,079 |
| Video135 | 30,373 | 489 | 4 | 387.48 | 1,283 |
| Video136 | 11,309 | 130 | 1 | 288.52 | 913 |
| Video137 | 27,926 | 609 | 20 | 193.32 | 688 |
| Video138 | 9,441 | 149 | 2 | 435.96 | 1,407 |
| Video139 | 3,003 | 45 | 2 | 355.68 | 1,371 |
| Video140 | 23,425 | 383 | 12 | 407.74 | 1,289 |
| Video141 | 4,281 | 64 | 5 | 281.2 | 845 |
| Video142 | 6,484 | 77 | 6 | 503.1 | 1,580 |
| Video143 | 7,755 | 96 | 3 | 494.94 | 1,238 |
| Video144 | 29,840 | 949 | 29 | 261.54 | 947 |
| Video145 | 15,094 | 155 | 2 | 386.1 | 1,476 |
| Video146 | 27,177 | 767 | 15 | 263.9 | 802 |
| Video147 | 4,380 | 55 | 11 | 436.84 | 1,286 |
| Video148 | 7,406 | 79 | 7 | 366.2 | 1,057 |
| Video149 | 45,243 | 1,081 | 8 | 496.54 | 1,325 |
| Video150 | 5,427 | 83 | 5 | 637.18 | 1,933 |
| Video151 | 3,728 | 68 | 3 | 322.96 | 1,117 |
| Video152 | 5,946 | 77 | 17 | 398 | 1,183 |
| Video153 | 4,967 | 62 | 32 | 476.94 | 1,478 |
| Video154 | 3,602 | 57 | 1 | 538.74 | 1,704 |
| Video155 | 48,492 | 1,370 | 51 | 287.72 | 960 |
| Video156 | 12,202 | 174 | 2 | 344.94 | 1,111 |
| Video157 | 5,526 | 133 | 83 | 228.08 | 748 |
| Video158 | 6,270 | 112 | 6 | 253.48 | 938 |
| Video159 | 2,845 | 38 | 6 | 413.06 | 1,216 |
| Video160 | 63,207 | 857 | 12 | 249.26 | 716 |
| Video161 | 7,500 | 132 | 2 | 274.76 | 991 |
| Video162 | 8,402 | 133 | 6 | 398.78 | 1,282 |
| Video163 | 5,349 | 40 | 1 | 291.72 | 953 |
| Video164 | 18,202 | 359 | 9 | 332.68 | 1,095 |
| Video165 | 57,586 | 1,669 | 57 | 324.9 | 1,255 |
| Video166 | 7,456 | 156 | 0 | 356.82 | 1,166 |
| Video167 | 5,010 | 65 | 3 | 325.88 | 1,274 |
| Video168 | 8,198 | 191 | 7 | 383.54 | 933 |
| Video169 | 13,455 | 228 | 6 | 304.58 | 1,290 |
| Video170 | 5,988 | 89 | 9 | 310.6 | 1,069 |
| Video171 | 35,639 | 1,197 | 105 | 391.54 | 1,501 |
| Video172 | 27,519 | 620 | 28 | 307.42 | 927 |
| Video173 | 14,627 | 190 | 4 | 292.32 | 879 |

(Continued)

TABLE 1 (Continued)

| Anonymized title | viewCount (Number of views) | likeCount (Number of likes) | commentCount (Number of comments) | Podcast Length (Length of podcast in seconds) | nWords (Number of words spoken in the podcast) |
|------------------|-----------------------------|-----------------------------|-----------------------------------|---|--|
| Video174 | 12,405 | 235 | 5 | 380.32 | 1,353 |
| Video175 | 44,666 | 920 | 4 | 315.42 | 974 |
| Video176 | 3,230 | 38 | 5 | 304.82 | 1,041 |
| Video177 | 20,920 | 508 | 1 | 364.84 | 1,509 |
| Video178 | 5,100 | 99 | 8 | 285.58 | 969 |
| Video179 | 2,456 | 25 | 1 | 317.22 | 927 |
| Video180 | 2,449 | 25 | 7 | 528.68 | 1782 |
| Video181 | 5,062 | 69 | 7 | 322.58 | 1,156 |
| Video182 | 18,034 | 304 | 2 | 349.04 | 1,069 |
| Video183 | 9,251 | 97 | 3 | 306.28 | 1,032 |
| Video184 | 63,037 | 2,124 | 1 | 311.3 | 1,182 |
| Video185 | 5,815 | 128 | 4 | 230.6 | 707 |
| Video186 | 36,746 | 533 | 99 | 441.42 | 1,255 |
| Video187 | 32,657 | 678 | 66 | 369.6 | 1,052 |
| Video188 | 30,926 | 469 | 1 | 341.4 | 1,217 |

as independent variables. The set of independent variables was further divided into affective linguistic and paralinguistic variables (or features).

Both affective linguistic and paralinguistic variables extracted from audio files typically include features related to the content, structure, or characteristics of speech or affective language within the audio. These variables can be extracted by employing techniques from speech processing, natural language processing (NLP), and audio signal processing.

In Table 3, the first column represents the average Affective Linguistic Variables, the second column represents whether the variable in the first column is Dynamic or Average, the third column represents Paralinguistic Variables, and the fourth column represents whether the variable in the third column is Average or Dynamic.

The average variables represent numeric values derived from the entire podcast duration. Dynamic variables are computed by segmenting the podcast into four equal parts and calculating the respective values for each segment. Dynamic variables are distinguished by the prefixes Q1, Q2, Q3, and Q4, which signify the statistical values pertaining to the first, second, third, and fourth segments of the podcast duration, respectively.

2.4 Data reduction management

The number of variables in comparison to the number of observations is very high. Typically, a dataset should contain 10 observations corresponding to a variable. For the sample used in this study, ideally, the number of rows should be at least 1,360, but we have only 188 observations in the sample. Therefore, we must reduce the number of variables to meet the standards.

We employ the Least Absolute Shrinkage Selection Operator (LASSO) regression method to reduce the number of variables. We then use a backward elimination procedure to further reduce the number of variables necessary to shape our final models.

Step 1: LASSO regression: This method is a regularization technique that enables us to select the most impactful variables while simultaneously reducing the number of parameters required to estimate. We estimate our linear regression model with the following expression:

$$\sum_{i=1}^n \left(y_i - \sum_j x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

where the λ is the penalty or shrinkage coefficient. We found that for our extensive dataset, a value of $\lambda = 20$ is appropriate and acceptable. With this shrinkage coefficient, we reduce the number of variables to approximately 50 remaining variables.

Step 2: Backwards elimination procedure: Given a model containing a large number of variables, we can apply the backward elimination procedure to further reduce the number of variables. This procedure means that we first estimate the full model with all the parameters and then select and eliminate the variable whose parameter estimate has the largest p -value. Once eliminated, we estimate the now smaller model and then repeat this elimination step again. We will continue until the model is sufficiently parsimonious. In this study, we use p -value = 0.20 as a stopping criterion.

Our present model formulation features a large number of interaction terms. In order to reduce redundancies in the interactions, we perform a step-wise elimination procedure to eliminate statistically insignificant parameter estimates of the interaction effects. We define the “given set of parameters” as the variables of the base mode along with the institutional conditions z and cultural dimension variables q .

The procedure is as follows:

- Estimate a multi-level model with the “given set of parameters.”
- Evaluate the p -values of the λ and δ parameter estimates.
- Select the variable with the largest p -value.
- The “given set of parameters” now excludes this selected variable.
- Repeat step 1.

TABLE 2 Independent variables and definitions.

| Sr. No. | Variable | Definition |
|---------|--------------------------------|--|
| 1 | podcastLength | Duration of the podcast episode in terms of time. |
| 2 | nWords | Total count of words spoken in the podcast episode. |
| 3 | SpeakLength_percent | Percentage of time spent speaking in the podcast episode. |
| 4 | posWords_percent | Percentage of positive sentiment words in the podcast episode. |
| 5 | negWords_percent | Percentage of negative sentiment words in the podcast episode. |
| 6 | pos_neg_ratio | Ratio of positive to negative sentiment words in the podcast episode. |
| 7 | TotalSentimentScore | Overall sentiment score of the podcast episode. |
| 8 | First30sec_TotalSentimentScore | Sentiment score of the first 30 s of the podcast episode. |
| 9 | First30sec_MinSentimentScore | Minimum sentiment score of the first 30 s of the podcast episode. |
| 10 | last30sec_TotalSentimentScore | Sentiment score of the last 30 s of the podcast episode. |
| 11 | First30sec_MinSentimentScore | Minimum sentiment score of the first 30 s of the podcast episode. |
| 12 | FirstQTotalSentimentScore | Total sentiment score of the first quarter duration of the podcast episode. |
| 13 | FirstQStandDevSentimentScore | Standard deviation of sentiment scores of the first quarter duration of the podcast episode. |
| 14 | SecondQTotalSentimentScore | Total sentiment score of the second quarter duration of the podcast episode. |
| 15 | SecondQStandDevSentimentScore | Standard deviation of sentiment scores of the second quarter duration of the podcast episode. |
| 16 | ThirdQTotalSentimentScore | Total sentiment score of the third quarter duration of the podcast episode. |
| 17 | ThirdQStandDevSentimentScore | Standard deviation of sentiment scores of the third quarter duration of the podcast episode. |
| 18 | FourthQTotalSentimentScore | Total sentiment score of the fourth quarter duration of the podcast episode. |
| 19 | FourthQStandDevSentimentScore | Standard deviation of sentiment scores of the fourth quarter duration of the podcast episode. |
| 20 | Timbre | Quality of sound determined by its harmonic content. |
| 21 | Number Syllables | Count of distinct speech units or sound combinations pronounced in the audio segment. |
| 22 | nPauses | Total count of breaks or silences in speech within the audio, typically indicating pauses or hesitations. |
| 23 | Articulation Rate | Speed at which speech sounds or syllables are pronounced within the audio recording. |
| 24 | Intonation | Variation in pitch and tone patterns used in speech to convey meaning or emotion. |
| 25 | Intonation Variability | Variation in pitch and tone patterns across the audio segment, indicating the flexibility and range of intonation. |
| 26 | Female | Gender identification of the speaker as female. |
| 27 | nPauses | Total count of breaks or silences in speech within the audio, typically indicating pauses or hesitations. |
| 28 | SilenceLength_percent | Percentage of time spent in silence within the audio segment. |
| 29 | Negative | Presence of negative sentiment in the audio segment. |
| 30 | Positive | Presence of positive sentiment in the audio segment. |
| 31 | Anger | Presence of anger sentiment in the audio segment. |
| 32 | Anticipation | Presence of anticipation sentiment in the audio segment. |
| 33 | Disgust | Presence of disgust sentiment in the audio segment. |
| 34 | Fear | Presence of fear sentiment in the audio segment. |
| 35 | Joy | Presence of joy sentiment in the audio segment. |
| 36 | Sadness | Presence of sadness sentiment in the audio segment. |
| 37 | Surprise | Presence of surprise sentiment in the audio segment. |
| 38 | Trust | Presence of trust sentiment in the audio segment. |
| 39 | Q1_positive | Presence of positive sentiment in the first quarter duration of the audio segment. |
| 40 | Q1_negative | Presence of negative sentiment in the first quarter duration of the audio segment. |
| 41 | Q1_anger | Presence of anger sentiment in the first quarter duration of the audio segment. |
| 42 | Q1_anticipation | Presence of anticipation sentiment in the first quarter duration of the audio segment. |
| 43 | Q1_disgust | Presence of disgust sentiment in the first quarter duration of the audio segment. |
| 44 | Q1_fear | Presence of fear sentiment in the first quarter duration of the audio segment. |
| 45 | Q1_joy | Presence of joy sentiment in the first quarter duration of the audio segment. |

(Continued)

TABLE 2 (Continued)

| Sr. No. | Variable | Definition |
|---------|-----------------|--|
| 46 | Q1_sadness | Presence of sadness sentiment in the first quarter duration of the audio segment. |
| 47 | Q1_surprise | Presence of surprise sentiment in the first quarter duration of the audio segment. |
| 48 | Q1_trust | Presence of trust sentiment in the first quarter duration of the audio segment. |
| 49 | Q2_positive | Presence of positive sentiment in the second quarter duration of the audio segment. |
| 50 | Q2_negative | Presence of negative sentiment in the second quarter duration of the audio segment. |
| 51 | Q2_anger | Presence of anger sentiment in the second quarter duration of the audio segment. |
| 52 | Q2_anticipation | Presence of anticipation sentiment in the second quarter duration of the audio segment. |
| 53 | Q2_disgust | Presence of disgust sentiment in the second quarter duration of the audio segment. |
| 54 | Q2_fear | Presence of fear sentiment in the second quarter duration of the audio segment. |
| 55 | Q2_joy | Presence of joy sentiment in the second quarter duration of the audio segment. |
| 56 | Q2_sadness | Presence of sadness sentiment in the second quarter duration of the audio segment. |
| 57 | Q2_surprise | Presence of surprise sentiment in the second quarter duration of the audio segment. |
| 58 | Q2_trust | Presence of trust sentiment in the second quarter duration of the audio segment. |
| 59 | Q3_positive | Presence of positive sentiment in the third quarter duration of the audio segment. |
| 60 | Q3_negative | Presence of negative sentiment in the third quarter duration of the audio segment. |
| 61 | Q3_anger | Presence of anger sentiment in the third quarter duration of the audio segment. |
| 62 | Q3_anticipation | Presence of anticipation sentiment in the third quarter duration of the audio segment. |
| 63 | Q3_disgust | Presence of disgust sentiment in the third quarter duration of the audio segment. |
| 64 | Q3_fear | Presence of fear sentiment in the third quarter duration of the audio segment. |
| 65 | Q3_joy | Presence of joy sentiment in the third quarter duration of the audio segment. |
| 66 | Q3_sadness | Presence of sadness sentiment in the third quarter duration of the audio segment. |
| 67 | Q3_surprise | Presence of surprise sentiment in the third quarter duration of the audio segment. |
| 68 | Q3_trust | Presence of trust sentiment in the third quarter duration of the audio segment. |
| 69 | Q4_positive | Presence of positive sentiment in the fourth quarter duration of the audio segment. |
| 70 | Q4_negative | Presence of negative sentiment in the fourth quarter duration of the audio segment. |
| 71 | Q4_anger | Presence of anger sentiment in the fourth quarter duration of the audio segment. |
| 72 | Q4_anticipation | Presence of anticipation sentiment in the fourth quarter duration of the audio segment. |
| 73 | Q4_disgust | Presence of disgust sentiment in the fourth quarter duration of the audio segment. |
| 74 | Q4_fear | Presence of fear sentiment in the fourth quarter duration of the audio segment. |
| 75 | Q4_joy | Presence of joy sentiment in the fourth quarter duration of the audio segment. |
| 76 | Q4_sadness | Presence of sadness sentiment in the fourth quarter duration of the audio segment. |
| 77 | Q4_surprise | Presence of surprise sentiment in the fourth quarter duration of the audio segment. |
| 78 | Q4_trust | Presence of trust sentiment in the fourth quarter duration of the audio segment. |
| 79 | Q1_nsyll | Count of distinct speech units or sound combinations pronounced in the first quarter duration of the audio segment. |
| 80 | Q1_npause | Total count of breaks or silences in speech within the first quarter duration of the audio segment. |
| 81 | Q1_sp_rate | The speed at which speech sounds or syllables are pronounced within the first quarter duration of the audio segment. |
| 82 | Q1_art_rate | The rate of articulation (speech clarity) within the first quarter duration of the audio segment. |
| 83 | Q1_sp_length | The total length of speech duration within the first quarter duration of the audio segment. |
| 84 | Q1_total_length | The total length of audio (including speech and silence) within the first quarter duration of the audio segment. |
| 85 | Q1_ratio | The ratio of certain audio feature(s) in the first quarter duration of the audio segment. |
| 86 | Q1_f0_mean | The mean fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 87 | Q1_f0_std | The standard deviation of fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 88 | Q1_f0_med | The median fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 89 | Q1_f0_min | The minimum fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 90 | Q1_f0_max | The maximum fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 91 | Q1_f0_q25 | The first quartile of fundamental frequency (pitch) in the first quarter duration of the audio segment. |

(Continued)

TABLE 2 (Continued)

| Sr. No. | Variable | Definition |
|---------|-----------------|---|
| 92 | Q1_f0_q75 | The third quartile of fundamental frequency (pitch) in the first quarter duration of the audio segment. |
| 93 | Q2_nsyll | Count of distinct speech units or sound combinations pronounced in the second quarter duration of the audio segment. |
| 94 | Q2_npause | Total count of breaks or silences in speech within the second quarter duration of the audio segment. |
| 95 | Q2_sp_rate | The speed at which speech sounds or syllables are pronounced within the second quarter duration of the audio segment. |
| 96 | Q2_art_rate | The rate of articulation (speech clarity) within the second quarter duration of the audio segment. |
| 97 | Q2_sp_length | The total length of speech duration within the second quarter duration of the audio segment. |
| 98 | Q2_total_length | The total length of audio (including speech and silence) within the second quarter duration of the audio segment. |
| 99 | Q2_ratio | The ratio of certain audio feature(s) in the second quarter duration of the audio segment. |
| 100 | Q2_ratio | The ratio of certain audio feature(s) in the second quarter duration of the audio segment. |
| 101 | Q2_f0_mean | The mean fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 102 | Q2_f0_std | The standard deviation of fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 103 | Q2_f0_med | The median fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 104 | Q2_f0_min | The minimum fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 105 | Q2_f0_max | The maximum fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 106 | Q2_f0_q25 | The first quartile of fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 107 | Q2_f0_Q75 | The third quartile of fundamental frequency (pitch) in the second quarter duration of the audio segment. |
| 108 | Q3_nsyll | Count of distinct speech units or sound combinations pronounced in the third quarter duration of the audio segment. |
| 109 | Q3_npause | Total count of breaks or silences in speech within the third quarter duration of the audio segment. |
| 110 | Q3_sp_rate | The speed at which speech sounds or syllables are pronounced within the third quarter duration of the audio segment. |
| 111 | Q3_art_rate | The rate of articulation (speech clarity) within the third quarter duration of the audio segment. |
| 112 | Q3_sp_length | The total length of speech duration within the third quarter duration of the audio segment. |
| 113 | Q3_total_length | The total length of audio (including speech and silence) within the third quarter duration of the audio segment. |
| 114 | Q3_ratio | The ratio of certain audio feature(s) in the third quarter duration of the audio segment. |
| 115 | Q3_f0_mean | The mean fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 116 | Q3_f0_std | The standard deviation of fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 117 | Q3_f0_med | The median fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 118 | Q3_f0_min | The minimum fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 119 | Q3_f0_max | The maximum fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 120 | Q3_f0_Q25 | The first quartile of fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 121 | Q3_f0_Q75 | The third quartile of fundamental frequency (pitch) in the third quarter duration of the audio segment. |
| 122 | Q3_joy | Presence of joy sentiment in the third quarter duration of the audio segment. |
| 123 | Q3_sadness | Presence of sadness sentiment in the third quarter duration of the audio segment. |
| 124 | Q3_surprise | Presence of surprise sentiment in the third quarter duration of the audio segment. |
| 125 | Q3_trust | Presence of trust sentiment in the third quarter duration of the audio segment. |
| 126 | Q4_positive | Presence of positive sentiment in the fourth quarter duration of the audio segment. |
| 127 | Q4_negative | Presence of negative sentiment in the fourth quarter duration of the audio segment. |
| 128 | Q4_anger | Presence of anger sentiment in the fourth quarter duration of the audio segment. |
| 129 | Q4_anticipation | Presence of anticipation sentiment in the fourth quarter duration of the audio segment. |
| 130 | Q4_disgust | Presence of disgust sentiment in the fourth quarter duration of the audio segment. |
| 131 | Q4_fear | Presence of fear sentiment in the fourth quarter duration of the audio segment. |
| 132 | Q4_joy | Presence of joy sentiment in the fourth quarter duration of the audio segment. |
| 133 | Q4_sadness | Presence of sadness sentiment in the fourth quarter duration of the audio segment. |
| 134 | Q4_surprise | Presence of surprise sentiment in the fourth quarter duration of the audio segment. |
| 135 | Q4_trust | Presence of trust sentiment in the fourth quarter duration of the audio segment. |
| 136 | Q4_nsyll | Count of distinct speech units or sound combinations pronounced in the fourth quarter duration of the audio segment. |
| 137 | Q4_npause | Total count of breaks or silences in speech within the fourth quarter duration of the audio segment. |

(Continued)

TABLE 2 (Continued)

| Sr. No. | Variable | Definition |
|---------|-----------------|---|
| 138 | Q4_sp_rate | The speed at which speech sounds or syllables are pronounced within the fourth quarter duration of the audio segment. |
| 139 | Q4_art_rate | The rate of articulation (speech clarity) within the fourth quarter duration of the audio segment. |
| 140 | Q4_sp_length | The total length of speech duration within the fourth quarter duration of the audio segment. |
| 141 | Q4_total_length | The total length of audio (including speech and silence) within the fourth quarter duration of the audio segment. |
| 142 | Q4_ratio | The ratio of certain audio feature(s) in the fourth quarter duration of the audio segment. |
| 143 | Q4_f0_mean | The mean fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 144 | Q4_f0_std | The standard deviation of fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 145 | Q4_f0_med | The median fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 146 | Q4_f0_min | The minimum fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 147 | Q4_f0_Q25 | The first quartile of fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 148 | Q4_f0_Q75 | The third quartile of fundamental frequency (pitch) in the fourth quarter duration of the audio segment. |
| 149 | Q1_prob | Probability of certain event or feature occurrence in the first quarter duration of the audio segment. |
| 150 | Q2_prob | Probability of certain event or feature occurrence in the second quarter duration of the audio segment. |
| 151 | Q3_prob | Probability of certain event or feature occurrence in the third quarter duration of the audio segment. |
| 152 | Q4_prob | Probability of certain event or feature occurrence in the fourth quarter duration of the audio segment. |

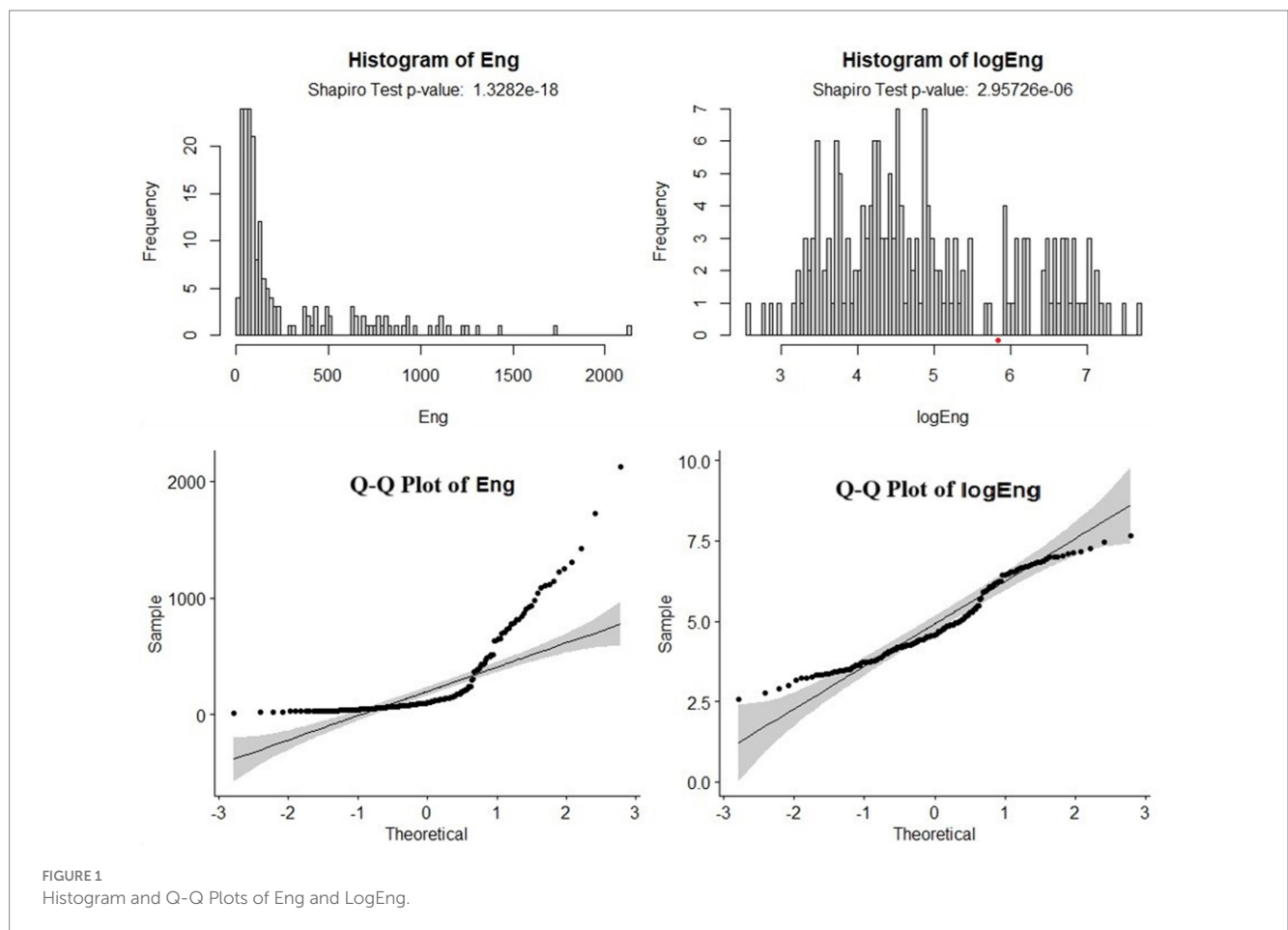


FIGURE 1 Histogram and Q-Q Plots of Eng and LogEng.

This procedure allows us to iterate through multiple model parameter estimations while reducing redundancies in the process. The final iteration is a parsimonious, interpretable model with statistically significant parameter estimates.

Figure 2 shows the calculated pairwise correlations between these variables. The affective linguistic variables show high correlations in the upper-left corner of the matrix. The dynamic paralinguistic features also show a highly correlated pattern across the four different

TABLE 3 Independent variables and variable types (average/dynamic).

| Sr. No. | Independent variable | Variable type | Sr. No. | Independent variable | Variable type |
|---------|--------------------------------|---------------|---------|----------------------|---------------|
| 1 | podcastLength | Average | 39 | Q1_positive | Dynamic |
| 2 | nWords | Average | 40 | Q1_negative | Dynamic |
| 3 | SpeakLength_percent | Average | 41 | Q1_anger | Dynamic |
| 4 | posWords_percent | Average | 42 | Q1_anticipation | Dynamic |
| 5 | negWords_percent | Average | 43 | Q1_disgust | Dynamic |
| 6 | pos_neg_ratio | Average | 44 | Q1_fear | Dynamic |
| 7 | TotalSentimentScore | Average | 45 | Q1_joy | Dynamic |
| 8 | First30sec_TotalSentimentScore | Average | 46 | Q1_sadness | Dynamic |
| 9 | First30sec_MinSentimentScore | Average | 47 | Q1_surprise | Dynamic |
| 10 | last30sec_TotalSentimentScore | Average | 48 | Q1_trust | Dynamic |
| 11 | First30sec_MinSentimentScore | Average | 49 | Q2_positive | Dynamic |
| 12 | FirstQTotalSentimentScore | Average | 50 | Q2_negative | Dynamic |
| 13 | FirstQStandDevSentimentScore | Average | 51 | Q2_anger | Dynamic |
| 14 | SecondQTotalSentimentScore | Average | 52 | Q2_anticipation | Dynamic |
| 15 | SecondQStandDevSentimentScore | Average | 53 | Q2_disgust | Dynamic |
| 16 | ThirdQTotalSentimentScore | Average | 54 | Q2_fear | Dynamic |
| 17 | ThirdQStandDevSentimentScore | Average | 55 | Q2_joy | Dynamic |
| 18 | FourthQTotalSentimentScore | Average | 56 | Q2_sadness | Dynamic |
| 19 | FourthQStandDevSentimentScore | Average | 57 | Q2_surprise | Dynamic |
| 20 | Timbre | Average | 58 | Q2_trust | Dynamic |
| 21 | Number Syllables | Average | 59 | Q3_positive | Dynamic |
| 22 | nPauses | Average | 60 | Q3_negative | Dynamic |
| 23 | Articulation Rate | Average | 61 | Q3_anger | Dynamic |
| 24 | Intonation | Average | 62 | Q3_anticipation | Dynamic |
| 25 | Intonation Variability | Average | 63 | Q3_disgust | Dynamic |
| 26 | Female | Average | 64 | Q3_fear | Dynamic |
| 27 | nPauses | Average | 65 | Q3_joy | Dynamic |
| 28 | SilenceLength_percent | Average | 66 | Q3_sadness | Dynamic |
| 29 | Negative | Average | 67 | Q3_surprise | Dynamic |
| 30 | Positive | Average | 68 | Q3_trust | Dynamic |
| 31 | Anger | Average | 69 | Q4_positive | Dynamic |
| 32 | Anticipation | Average | 70 | Q4_negative | Dynamic |
| 33 | Disgust | Average | 71 | Q4_anger | Dynamic |
| 34 | Fear | Average | 72 | Q4_anticipation | Dynamic |
| 35 | Joy | Average | 73 | Q4_disgust | Dynamic |
| 36 | Sadness | Average | 74 | Q4_fear | Dynamic |
| 37 | Surprise | Average | 75 | Q4_joy | Dynamic |
| 38 | Trust | Average | 76 | Q4_sadness | Dynamic |
| | | | 77 | Q4_surprise | Dynamic |
| | | | 78 | Q4_trust | Dynamic |
| | | | 79 | Q1_nsyll | Dynamic |
| | | | 80 | Q1_npause | Dynamic |
| | | | 81 | Q1_sp_rate | Dynamic |
| | | | 82 | Q1_art_rate | Dynamic |
| | | | 83 | Q1_sp_length | Dynamic |

(Continued)

TABLE 3 (Continued)

| Sr. No. | Independent variable | Variable type | Sr. No. | Independent variable | Variable type |
|---------|----------------------|---------------|---------|----------------------|---------------|
| | | | 84 | Q1_total_length | Dynamic |
| | | | 85 | Q1_ratio | Dynamic |
| | | | 86 | Q1_f0_mean | Dynamic |
| | | | 87 | Q1_f0_std | Dynamic |
| | | | 88 | Q1_f0_med | Dynamic |
| | | | 89 | Q1_f0_min | Dynamic |
| | | | 90 | Q1_f0_max | Dynamic |
| | | | 91 | Q1_f0_q25 | Dynamic |
| | | | 92 | Q1_f0_q75 | Dynamic |
| | | | 93 | Q2_nsyll | Dynamic |
| | | | 94 | Q2_npause | Dynamic |
| | | | 95 | Q2_sp_rate | Dynamic |
| | | | 97 | Q2_art_rate | Dynamic |
| | | | 98 | Q2_sp_length | Dynamic |
| | | | 99 | Q2_total_length | Dynamic |
| | | | 100 | Q2_ratio | Dynamic |
| | | | 101 | Q2_f0_mean | Dynamic |
| | | | 102 | Q2_f0_std | Dynamic |
| | | | 103 | Q2_f0_med | Dynamic |
| | | | 104 | Q2_f0_min | Dynamic |
| | | | 105 | Q2_f0_max | Dynamic |
| | | | 106 | Q2_f0_q25 | Dynamic |
| | | | 107 | Q2_f0_Q75 | Dynamic |
| | | | 108 | Q3_nsyll | Dynamic |
| | | | 109 | Q3_npause | Dynamic |
| | | | 110 | Q3_sp_rate | Dynamic |
| | | | 111 | Q3_art_rate | Dynamic |
| | | | 112 | Q3_sp_length | Dynamic |
| | | | 113 | Q3_total_length | Dynamic |
| | | | 114 | Q3_ratio | Dynamic |
| | | | 115 | Q3_f0_mean | Dynamic |
| | | | 116 | Q3_f0_std | Dynamic |
| | | | 117 | Q3_f0_med | Dynamic |
| | | | 118 | Q3_f0_min | Dynamic |
| | | | 119 | Q3_f0_max | Dynamic |
| | | | 120 | Q3_f0_Q25 | Dynamic |
| | | | 121 | Q3_f0_Q75 | Dynamic |
| | | | 122 | Q3_joy | Dynamic |
| | | | 123 | Q3_sadness | Dynamic |
| | | | 124 | Q3_surprise | Dynamic |
| | | | 125 | Q3_trust | Dynamic |
| | | | 126 | Q4_positive | Dynamic |
| | | | 127 | Q4_negative | Dynamic |
| | | | 128 | Q4_anger | Dynamic |

(Continued)

TABLE 3 (Continued)

| Sr. No. | Independent variable | Variable type | Sr. No. | Independent variable | Variable type |
|---------|----------------------|---------------|---------|----------------------|---------------|
| | | | 129 | Q4_anticipation | Dynamic |
| | | | 130 | Q4_disgust | Dynamic |
| | | | 131 | Q4_fear | Dynamic |
| | | | 132 | Q4_joy | Dynamic |
| | | | 133 | Q4_sadness | Dynamic |
| | | | 134 | Q4_surprise | Dynamic |
| | | | 135 | Q4_trust | Dynamic |
| | | | 136 | Q4_nsyll | Dynamic |
| | | | 137 | Q4_npause | Dynamic |
| | | | 138 | Q4_sp_rate | Dynamic |
| | | | 139 | Q4_art_rate | Dynamic |
| | | | 140 | Q4_sp_length | Dynamic |
| | | | 141 | Q4_total_length | Dynamic |
| | | | 142 | Q4_ratio | Dynamic |
| | | | 143 | Q4_f0_mean | Dynamic |
| | | | 144 | Q4_f0_std | Dynamic |
| | | | 145 | Q4_f0_med | Dynamic |
| | | | 146 | Q4_f0_min | Dynamic |
| | | | 147 | Q4_f0_Q25 | Dynamic |
| | | | 148 | Q4_f0_Q75 | Dynamic |
| | | | 149 | Q1_prob | Dynamic |
| | | | 150 | Q2_prob | Dynamic |
| | | | 151 | Q3_prob | Dynamic |
| | | | 152 | Q4_prob | Dynamic |
| | | | 153 | | |

quarters of the podcast. Importantly, the upper-right and lower-left squares of the matrix show the correlations between the affective linguistic and paralinguistic features. The matrix values suggest that these features are moderately correlated with each other.

3 Results

This chapter presents the results of a stepwise machine learning model approach, starting with average affective linguistic and paralinguistic features and gradually adding both dynamic affective linguistic and paralinguistic features, as well as the interaction between these features, to substantiate Conjecture 1. Subsequently, we conclude with two distinct models that separate the female and male podcast presenters, thus addressing Conjecture 2.

3.1 Average affective linguistics and paralinguistic models

Model 1: The average affective linguistics model was constructed by incorporating variables such as podcast length, gender, and affective linguistic features as predictors. The adjusted R^2 was 0.06 with a p -value of 0.02. Focusing on the most significant regression coefficients, Table 4

and Figure 3 show that especially fear (coefficient: $-0.42, p > 0.02$) and total sentiment scores (coefficient: $-0.49, p > 0.01$) had negative effects on audience engagement. Word count (coefficient: $0.34, p > 0.12$) and anticipation (coefficient: $0.29, p > 0.06$) had positive effects.

In Model 2, we integrate Average Linguistics and Average Paralinguistic Features together, as shown in Table 5 and Figure 4. There was no significant rise in the explained variance, as predictors explained an adjusted R^2 of 0.07 with a p -value of 0.00. Most apparent is that in model 2, word count (coefficient: $0.68, p > 0.02$) has a positive impact on engagement, and fear (coefficient: $-0.46, p > 0.01$) as well as total sentiment score (coefficient: $-0.55, p > 0.00$) have negative effects on engagement. However, more paralinguistic features become salient in the model. Especially timbre (coefficient: $1.15, p > 0.08$) has a positive effect, and intonation q25 (coefficient: $-0.81, p > 0.19$) has a negative effect.

The conclusion from model 2 is that (a) not much explained variance is added compared to model 1, but that (b) some average paralinguistic features become significant variables.

3.2 Dynamic affective linguistics and paralinguistic effects

Model 3: Affective Linguistics is dynamic, which includes both the average or general affective linguistic features and the affective

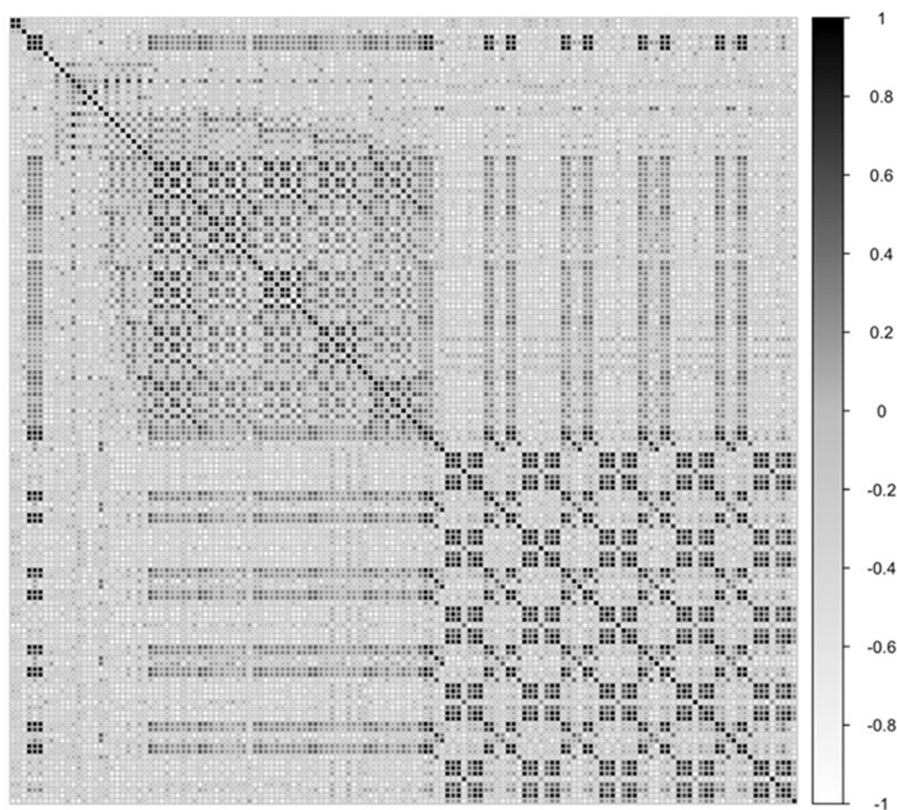


FIGURE 2
Correlation plot of independent variables.

TABLE 4 Model 1 linguistic average.

| Variable | Coefficients | <i>p</i> -value |
|---------------------------|--------------|-----------------|
| Intercept | 0 | 0.98 |
| Word Count | 0.34 | 0.15 |
| Anticipation | 0.29 | 0.06 |
| Surprise | 0.27 | 0.05 |
| Positive Words Percentage | 0.24 | 0.09 |
| Gender | -0.01 | 0.91 |
| Podcast Length | -0.35 | 0.12 |
| Fear | -0.42 | 0.02 |
| Total Sentiment Score | -0.49 | 0.01 |

linguistic features broken down into quarter-wise segments (dynamic affective linguistic features), explains more variance compared to model 2, as it achieved an adjusted R^2 of 0.22 with a p -value of 0.00. This model in Table 6 and Figure 5 shows a wider variety of predictive variables explaining audience engagement. Especially salient is that disgust (coefficient: 3.05, $p > 0.03$) has a positive effect on engagement, while disgust at the more granular level shows negative effects on engagement. In particular, disgust Q4 (coefficient: -0.94, $p > 0.01$), disgust Q1 (coefficient: -1.01, $p > 0.03$) and disgust Q2 (coefficient: -1.04, $p > 0.04$) exhibit negative effects. Similarly, fear, in general, has an especially negative effect on engagement (coefficient: -4.16, $p > 0.00$), while fear at the more granular level

shows negative effects on engagement. Here, fear Q2 (coefficient: 1.52, $p > 0.00$), fear Q1 (coefficient: 1.35, $p > 0.00$), and fear Q3 (coefficient: 1.03, $p > 0.00$) exhibit negative effects. It should be noted that the average negative (coefficient: 1.17, $p > 0.00$) has a positive effect on engagement, whereas negative Q2 (coefficient: -0.82, $p > 0.02$) has a negative effect.

Model 4, which adds the Average and Dynamic Affective Linguistics and Paralinguistic features to the podcast, explains more variance than models 1, 2, and 3, as it achieved an adjusted R^2 of 0.26 with a p -value of 0.00. The model shown in Table 7 and Figure 6 illustrates yet another different variety in predictive variables explaining audience engagement that are predominantly dynamic paralinguistic features. Especially salient is that timbre Q3 (coefficient: 6.01, $p > 0.01$), timbre Q2 (coefficient: 3.65, $p > 0.05$), intonation Q1 q75 (coefficient: 2.91, $p > 0.05$), the number of syllables (coefficient: 2.91, $p > 0.00$), and speak length (coefficient: 1.43, $p > 0.02$) have positive effects on engagement. It is noteworthy that anticipation in general (coefficient: 1.49, $p > 0.00$), just like in the earlier models, has a positive effect on audience engagement. On the other hand, especially the more granular paralinguistic features, such as speak length Q3 (coefficient: -1.62, $p > 0.01$), a number of syllables Q1 (coefficient: -2.52, $p > 0.00$), Q2 intonation p75 (coefficient: -2.74, $p > 0.07$), Q3 intonation q25 (coefficient: -3.03, $p > 0.00$), Q1 timbre (coefficient: -3.10, $p > 0.07$), and Q1 intonation q75 (coefficient: -3.72, $p > 0.02$) show negative effects on engagement. It should be noted that only one general marker, the number of pauses (coefficient: -1.85, $p > 0.01$), has a negative effect on engagement.

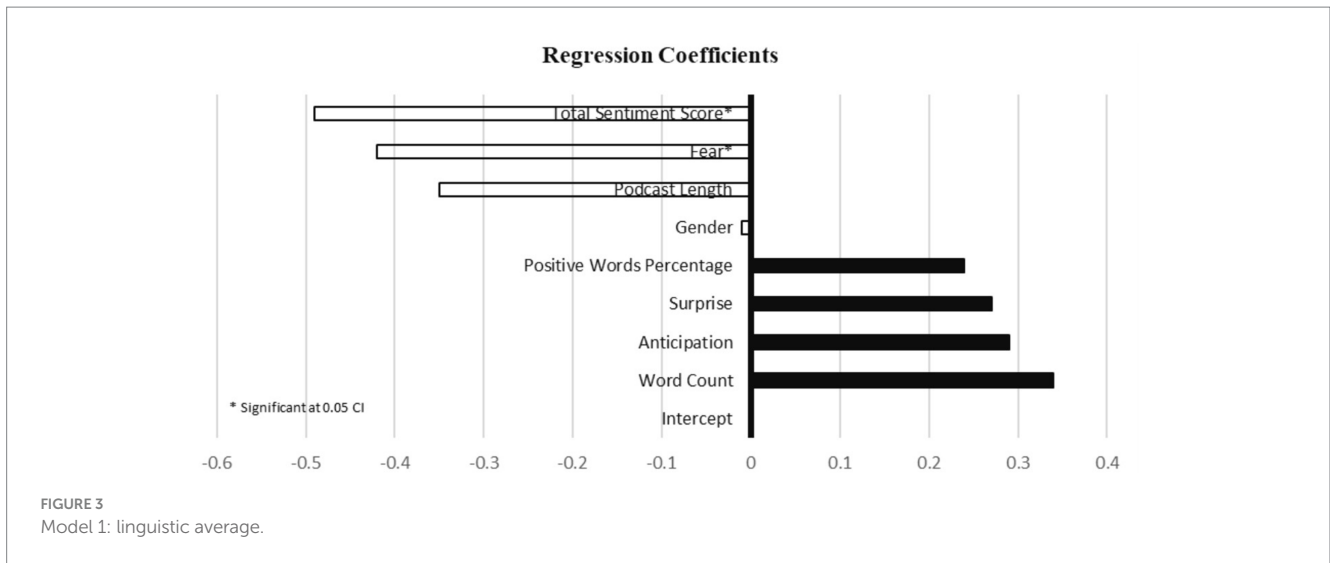


FIGURE 3 Model 1: linguistic average.

TABLE 5 Model 2 linguistic and paralinguistic average.

| Variable | Coefficients | p-value |
|---------------------------|--------------|---------|
| Intercept | 0.00 | 0.92 |
| Timbre | 1.15 | 0.07 |
| Word Count | 0.68 | 0.02 |
| Anticipation | 0.30 | 0.05 |
| Positive Words Percentage | 0.27 | 0.05 |
| Surprise | 0.26 | 0.06 |
| Gender | -0.07 | 0.55 |
| Max. Intonation | -0.22 | 0.15 |
| Articulation Rate | -0.24 | 0.13 |
| Podcast Length | -0.30 | 0.33 |
| Number of Pauses | -0.36 | 0.15 |
| Fear | -0.46 | 0.13 |
| Total Sentiment Score | -0.55 | 0.01 |
| Intonation q25 | -0.81 | 0.19 |

Conclusion about models 3 and 4: (a) as expected, the models explain more variance, and (b) it is quite interesting that dynamic or granular paralinguistic features become the predominant predictive features of audience engagement.

3.3 Dynamic affective linguistics and paralinguistic interaction effects

In model 5: Affective Linguistics and Paralinguistics: Dynamic interaction effects, including interaction terms allows for a more advanced analysis of how affective linguistic and paralinguistic features interact to influence audience engagement. Here more variance is explained compared to the earlier models 1, 2, 3, and 4. The adjusted R^2 is 0.38 with a p -value of 0.00. Surprisingly, as shown in Table 8 and Figure 7, there is another set of predictive features or variables that explain audience engagement in a positive or negative

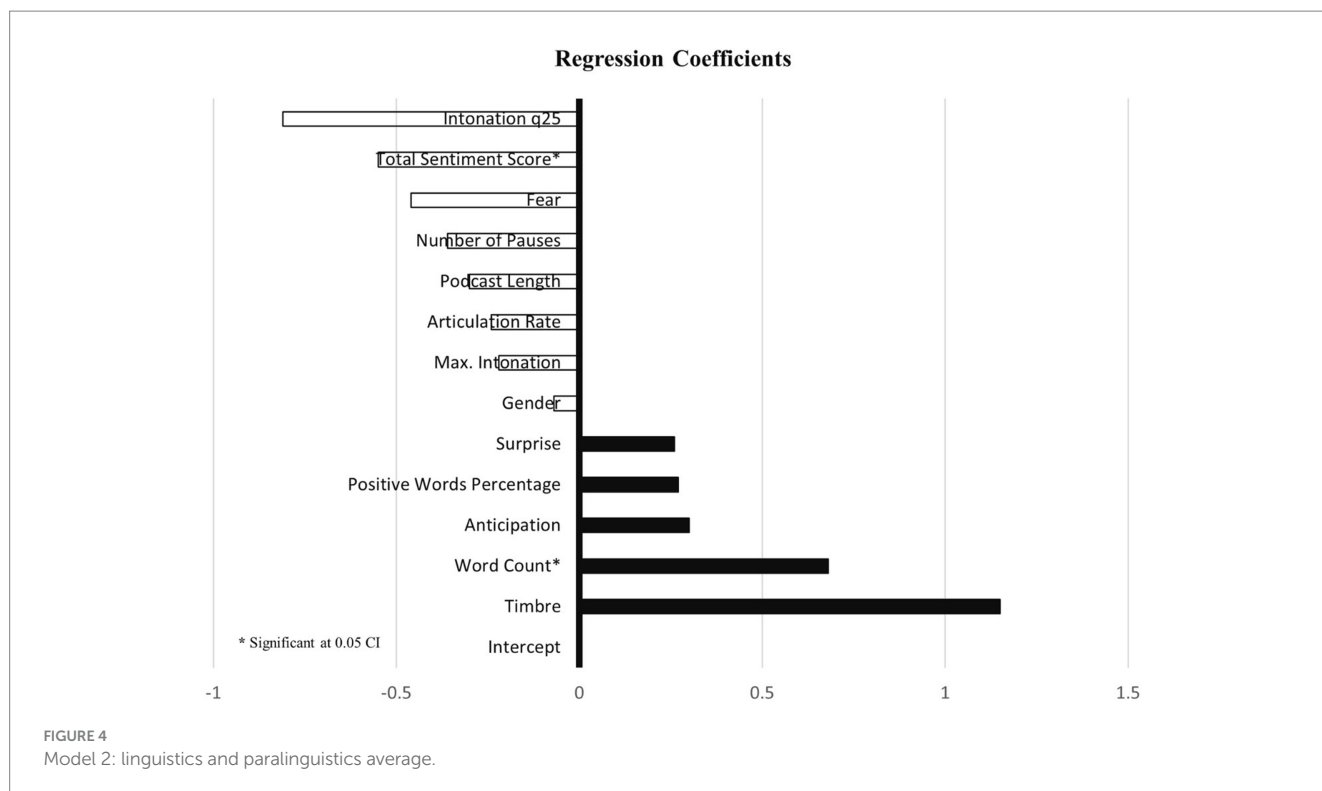
way. First in predicting audience engagement, the interaction between Q1 surprise x Q1 intonation q75 is most significant (coefficient: 4.76, $p > 0.00$), and to a somewhat lesser extent, Q4 surprise x Q4 articulation rate (coefficient: 0.33, $p > 0.00$), Q1 negative x Q1 intonation q25 (coefficient: 0.29, $p > 0.00$), Q4 positive x Q4 articulation rate (coefficient: 0.28, $p > 0.00$), and Q4 anticipation x Q4 articulation rate (coefficient: 0.21, $p > 0.04$) are positively affecting audience engagement. However, the average or general paralinguistic feature, the number of syllables (coefficient: 2.63, $p > 0.01$) also positively affects engagement. Do the general affective linguistic features anticipation (coefficient: 0.51, $p > 0.00$) and surprise (coefficient: 0.34, $p > 0.00$) have positive effects? The positive effect of podcast length intonation is Q1 (coefficient: 1.55, $p > 0.07$).

There are also many features or variables with negative effects on engagement, such as the interaction effects of Q1 surprise x Q1 timbre (coefficient: -4.05, $p > 0.00$), Q1 fear x Q1 min intonation (coefficient: -0.37, $p > 0.00$), and Q1 surprise x Q1 intonation variability (coefficient: -0.97, $p > 0.00$). Q1 number of syllables (coefficient: -1.41, $p > 0.00$) and Q3 number of syllables (coefficient: -1.52, $p > 0.00$) and as well as the number of pauses (coefficient: -1.54, $p > 0.04$) should be taken note of.

The conclusion of model 5 is that a myriad of affective linguistic and paralinguistic features affect the listener’s engagement in both positive and negative ways. This adds to conjecture 1 that by adding interaction effects of affective linguistic and paralinguistic markers, there is more variance in audience engagement that can be explained.

3.4 Female and male prediction models

Model 6: Table 9 and Figure 8 show the coefficients and corresponding p -values of the female podcast presenters Linguistics & Paralinguistics—Dynamic Interaction Effects. By building this model, we notice in Table 9 that a substantially higher variance compared to the earlier models 1, 2, 3, 4, and 5 is achieved as it shows an adjusted R^2 of 0.50 with a p -value of 0.00. This added variance by taking a gender dimorphic perspective is indeed substantial. We seek to summarize the most important findings from this model. First is that most positive variables entail granular interactions between both positive and more negative emotions with paralinguistic features, while only one granular



feature has a positive effect, namely Q4 anticipation, which has a coefficient of 0.34, $p > 0.01$. Obvious are the negative granular features Q3 anger x Q3 timbre (coefficient: 1.87, $p > 0.04$), Q3 disgust x Q3 intonation q25 (coefficient: 1.55, $p > 0.04$), and negative Q3 x Q3 articulation rate (coefficient: 0.68, $p > 0.00$). Yet the more granular positive emotions with a specific granular speak characteristic also have positive effects: Q1 anticipation x Q1 intonation variability (coefficient: 0.42, $p > 0.01$) and Q4 surprise x Q4 articulation rate (coefficient: 0.37, $p > 0.04$).

Where the negative effect variables are concerned, a similar trend can be observed. First, Q3 anger x Q3 intonation (coefficient: -1.49, $p > 0.11$), Q1 anger x Q1 timbre (coefficient: -0.38, $p > 0.01$), and Q3 disgust x Q3 timbre (coefficient: -1.34, $p > 0.01$). Other granular features show negative interaction effects too: Q4 surprise x Q4 min intonation q25 (coefficient: -0.37, $p > 0.02$), Q4 surprise x Q4 minimum intonation (coefficient: -0.37, $p > 0.03$), Q3 surprise x Q3 articulation rate (coefficient: -0.25, $p > 0.01$), Q4 positive x Q4 max intonation (coefficient: -0.29, $p > 0.04$), and Q2 joy x Q2 timbre (coefficient: -0.22, $p > 0.06$). Note also that podcast length, just like in other models, has a negative effect (coefficient: -0.36, $p > 0.01$).

Model 7: Male Speaker Affective Linguistics & Paralinguistics: Dynamic Interaction Effects explores the intricate dynamics of language processing of the male podcast presenters. By creating this machine learning model, we notice in Table 10 and Figure 9 that the variance is more explained compared to the earlier models 1, 2, 3, 4, and 5, as it achieved an adjusted R^2 of 0.73 with a p -value of 0.00. This shows a substantial added variance by taking a gender dimorphic perspective; males, in this case, are vast. But we also notice that many variables that relate to the interaction of affective linguistic and paralinguistic features at the granular level play a role as well as expected as well as the general or average level variables predict both engagement in positive and negative ways. We seek to summarize the most important findings from this model.

As the positive effect of the average features is concerned, we notice positive effects of the number of pauses (coefficient: 1.21, $p > 0.00$) and number of syllables (coefficient: 0.99, $p > 0.07$), and for the negative effects of general features, we notice word count (coefficient: -0.92, $p > 0.01$) and podcast length (coefficient: -1.00, $p > 0.05$). For dynamic interaction effects, we notice Q2 disgust x Q2 articulation rate (coefficient: -3.42, $p > 0.03$), Q3 surprise x Q3 articulation rate (coefficient: -0.93, $p > 0.00$), and Q4 sadness x Q4 max intonation (coefficient: -0.86, $p > 0.00$). Note that podcast length (coefficient: -1.00, $p > 0.05$) has had a negative effect as in many models.

As far as the positive effect of the interaction effects of the dynamic features is concerned, we notice somewhat similar effects as in the female podcast group (model 6). Here the most obvious positive effects are: Q2 negative x Q2 articulation rate (coefficient: 1.49, $p > 0.00$), Q3 anger x Q3 articulation rate (coefficient: 1.11, $p > 0.00$), Q3 surprise x Q3 max intonation (coefficient: 0.64, $p > 0.00$), Q3 anticipation x Q3 articulation rate (coefficient: 0.64, $p > 0.00$), and Q3 anticipation x Q3 timbre (coefficient: 0.58, $p > 0.00$). In short, both the negative and positive emotional words amplified by speech characteristics matter positively in predictive engagement.

For the negative effects of the interaction effects of the dynamic variables, we notice somewhat similar effects as in the female podcast group (model 6). The most obvious effects are: Q2 disgust x Q2 articulation rate (coefficient: -3.43, $p > 0.00$), Q3 surprise x Q3 articulation rate (coefficient: -0.93, $p > 0.00$), Q4 sadness x Q4 maximum intonation (coefficient: -0.86, $p > 0.00$), and Q2 disgust x Q2 articulation rate (coefficient: -0.55, $p > 0.06$). Here too, both the negative and positive emotional words amplified by speech characteristics matter.

Let us stock models 6 and 7: by segmenting the podcasts by female and male podcast presenters, a greater amount of variance is explained, and most obvious are that both average variables predict

TABLE 6 Model 3 linguistic dynamic.

| Variable | Coefficients | p-value |
|---------------------------------|--------------|---------|
| Intercept | -0.00 | 0.99 |
| Disgust | 3.05 | 0.03 |
| Q2 Fear | 1.52 | 0.00 |
| Q1 Fear | 1.35 | 0.00 |
| Anticipation | 1.27 | 0.00 |
| Negative | 1.17 | 0.00 |
| Q3 Fear | 1.03 | 0.00 |
| Q4 Fear | 0.99 | 0.02 |
| Q3 Surprise | 0.42 | 0.00 |
| Q2 Sadness | 0.39 | 0.07 |
| Last 30 s Sent. Score | 0.35 | 0.01 |
| Joy | 0.34 | 0.14 |
| Speak Length | 0.23 | 0.04 |
| Gender | 0.08 | 0.48 |
| 1 st 30 s Sent Score | -0.22 | 0.09 |
| Q4 Anger | -0.23 | 0.16 |
| Q4 Std. dev. Sent Score | -0.30 | 0.03 |
| Q1 Joy | -0.34 | 0.08 |
| Podcast Length | -0.39 | 0.01 |
| Q3 Sadness | -0.46 | 0.02 |
| Q2 Anticipation | -0.46 | 0.02 |
| Q1 Anticipation | -0.54 | 0.01 |
| Q3 Anticipation | -0.66 | 0.00 |
| Q3 Disgust | -0.71 | 0.11 |
| Q2 Negative | -0.82 | 0.02 |
| Q4 Disgust | -0.94 | 0.01 |
| Q1 Disgust | -1.01 | 0.03 |
| Q2 Disgust | -1.04 | 0.04 |
| Fear | -4.16 | 0.00 |

audience engagement as well as the granular or dynamic variations where affective words and paralinguistic cues combined evoke both positive and negative effects in audience engagement. These latter two models substantiate conjecture 2 that by designing gender-dimorph machine learning models more variance is explained.

Table 11 summarizes the overall statistics of the different models and indeed shows according to the conjectures made that: (1) average affective linguistic and paralinguistic features show low predictive value; (2) by adding the dynamic interaction effects of affective linguistic and paralinguistic features, more variance in audience engagement is explained; and (3) when segmenting by the gender of the podcast presenter, even more variance is explained.

4 Discussion

Since podcasting has emerged as a significant oral medium for thought leaders and influencers to disseminate their ideas and insights

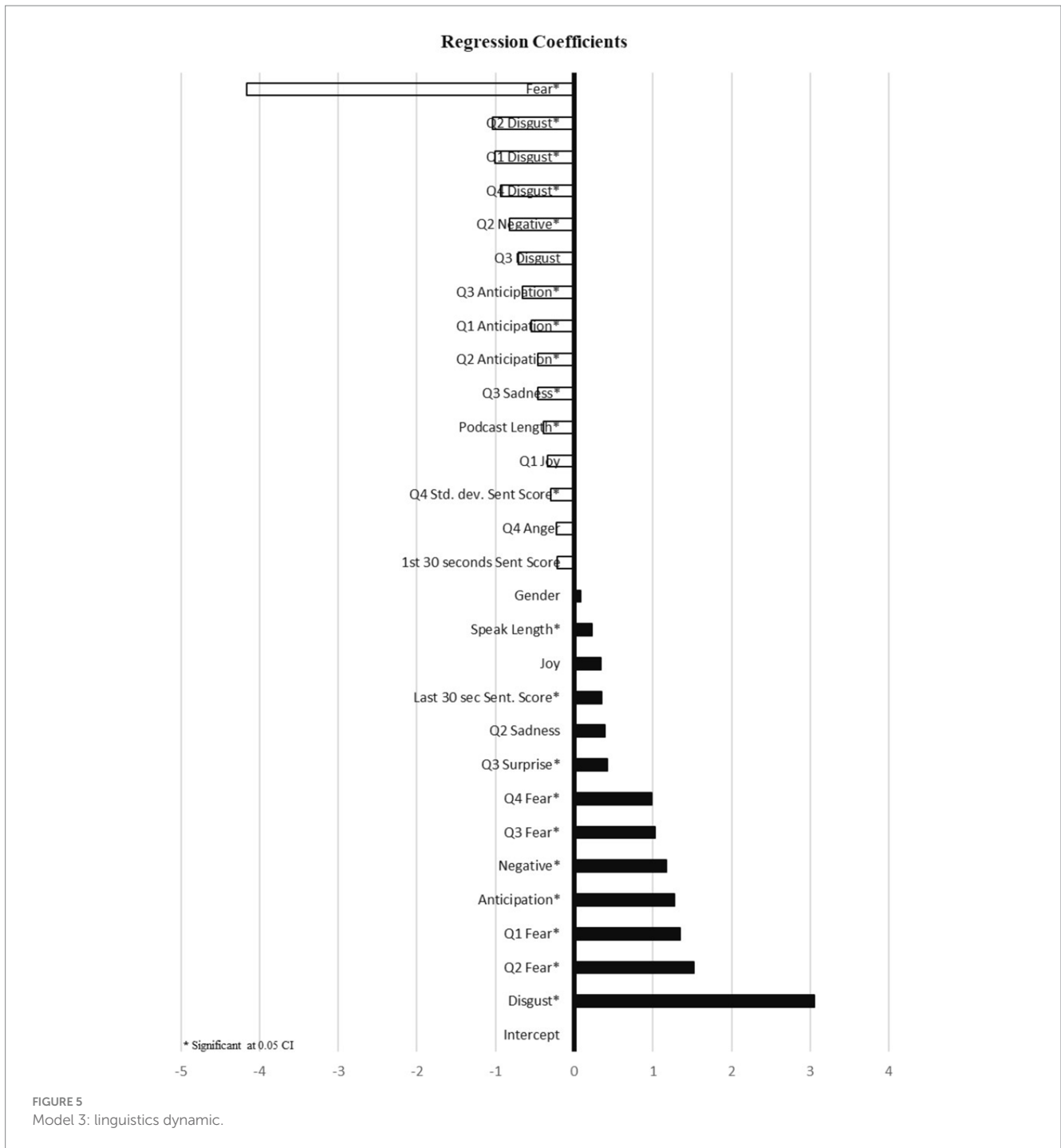
to a broader audience, it is crucial to investigate which affective linguistic and paralinguistic features drive audience engagement in podcasts. In other words, it is not only important what podcast presenters say but also how words are said. A machine learning approach allows how a large bundle of affective linguistic and paralinguistic features predicts engagement.

To discuss the findings of this specific study, we conceptualize podcast presenters as thought leaders meticulously chosen by the TEDx organization. These presenters are allotted a predetermined amount of time to share their most pertinent and surprising insights, a phenomenon often referred to as the “salesperson effect” (Baek and Falk, 2018). In simpler terms, the audience listening to the podcasts typically inquires, “What is new in this program?” and “What can I, in turn, share with my social network members?” (see also Perks et al., 2019). Based on the study of Van Zant and Berger (2020, p. 661), who noted that individuals modulate their voices during persuasion attempts in order to enhance their persuasiveness, this study examines the “performance persona.” This persona plays a dominant role during interactions with an attentive, albeit typically silent, audience, as described by Berry and Brown (2019, p. 14), where a solo speaker engages with an attentive audience. Here, podcast presenters aim to amplify and express their expertise by presenting themselves as confident and competent—a perception encouraged by both the TEDx organizers and the audience, thereby facilitating effective communication (Tsagkias et al., 2010).

It should be noted that this study primarily focused on the emotional or affective linguistic features and not the overall linguistic features of podcasts that evoke emotions, thereby fostering an emotional connection or contagion with the audience (Rocklage et al., 2018). Therefore, emotional features, such as anticipation, surprise, disgust, fear, joy, and sadness, were considered (Liu et al., 2022). It should also be noted that when we deal with the interactions between affective linguistic and paralinguistic features, they are commonly referred to as emotional prosody (Lausen and Schacht, 2018).

Despite the findings in most studies that focus on persuasion effects of average markers, we posited that the inclusion of average affective linguistic and paralinguistic features in a predictive model would not significantly enhance the accuracy of audience engagement for the podcast, and indeed our analysis, conducted within the context of a TEDx podcast, confirmed this intuition. Specifically, the average affective linguistic and paralinguistic features accounted for a minimal amount of variance, with an adjusted R^2 of 0.06 in model 1 and an adjusted R^2 of 0.07 in model 2. Several subsequent models demonstrated considerably better predictive performance, which we discuss in the following sections.

We examined the dynamic affective linguistic and paralinguistic features based on the underlying assumption that these TEDx podcasts function as narratives with a beginning, middle, and end. This narrative structure potentially enhances listener interest and persuasiveness while also preventing monotony (Van Krieken et al., 2015) and adding emotional volatility to the podcast (Berger et al., 2021). As suggested by Dowling and Miller (2019), podcasts can be viewed as cultural products of public radio and immersive online storytelling, which often employ narrative techniques. Similarly to musical compositions, podcasts adhere to a “podcast logic” and are consumed linearly, with a structured beginning, middle, and end. This helps to maintain listener engagement (adapted from Lindgren, 2023, p. 704). In literary terms, this narrative structure is frequently referred



to as a plot, and many podcast presenters employ scripted storytelling (Sotério and Linhares Queiroz, 2023). It is important to note that our identification of a “plot structure” is an interpretive assumption based on our being “participant listeners,” so to state, while listening to a sample TEDx podcast used in the study. However, we employed a coarse-grained approach by segmenting the podcasts into equal-time episodes to investigate the temporal disparities of the affective linguistic and paralinguistic features. These features, which we have termed dynamic prosodic contours or episodes, were analyzed to discern nuanced patterns or emotional volatility. These contours specifically accentuate certain affective linguistic or paralinguistic

dimensions, for instance, at the beginning (e.g., Q1) or toward the end (e.g., Q4).

First, supporting Conjecture 1, the inclusion of dynamic or granular affective linguistic and paralinguistic features in prediction model 3 explained additional variance, achieving an adjusted R^2 of 0.22, and prediction model 4 explained additional variance, achieving an adjusted R^2 of 0.26. These models represent an increase of approximately 20% in explained variance when compared to the first two models. Second, also in support of Conjecture 1, incorporating interactions between the dynamic affective linguistic and paralinguistic features in Model 5 further increased the explained

TABLE 7 Model 4 linguistic and paralinguistic dynamic.

| Variable | Coefficients | p-value |
|-------------------------------------|--------------|---------|
| Intercept | -0.01 | 0.88 |
| Q3 Timbre | 6.01 | 0.01 |
| Q2 Timbre | 3.65 | 0.05 |
| Q1 Intonation q75 | 2.91 | 0.05 |
| Number of Syllables | 2.91 | 0.00 |
| Anticipation | 1.49 | 0.00 |
| Q1 Speak Length | 1.43 | 0.02 |
| Q4 Number Pauses | 0.83 | 0.04 |
| Q2 Number Pauses | 0.63 | 0.14 |
| Q1 Max Intonation | 0.60 | 0.00 |
| Joy | 0.35 | 0.11 |
| Q3 Surprise | 0.34 | 0.01 |
| Q4 Anticipation | 0.28 | 0.04 |
| Last 30 Seconds Sentiment | 0.25 | 0.02 |
| Gender | 0.03 | 0.80 |
| Podcast Length | -0.14 | 0.78 |
| 1 st 30 s Min. Sentiment | -0.20 | 0.08 |
| Q4 Disgust | -0.31 | 0.01 |
| 1 st 30 Second Sentiment | -0.35 | 0.01 |
| Q1 Joy | -0.45 | 0.02 |
| Articulation Rate | -0.46 | 0.02 |
| Q2 Anticipation | -0.59 | 0.00 |
| Q1 Anticipation | -0.62 | 0.00 |
| Max Intonation | -0.63 | 0.00 |
| Q3 Anticipation | -0.67 | 0.00 |
| Q3 Speak Length | -1.62 | 0.00 |
| Number Pauses | -1.85 | 0.01 |
| Q1 Number Syllables | -2.52 | 0.00 |
| Q2 Intonation q75 | -2.74 | 0.07 |
| Q3 Intonation q25 | -3.03 | 0.01 |
| Q1 Timbre | -3.10 | 0.07 |
| Q3 Intonation q75 | -3.72 | 0.02 |

variance, resulting in an adjusted R^2 of 0.38, which is 10% more than that achieved by Model 4. Thus, model 5 provides additional evidence that the interactions between linguistic expressions and prosodic contours contribute significantly to variance explanation. Most conspicuous are the dynamic, granular affective linguistic and paralinguistic interactions associated with surprise and anticipation, which manifest both positive and negative effects on audience engagement, as well as the average affective linguistic features of surprise and anticipation.

It should be noted that in the model, additional average variables about the podcast, such as podcast length, average number of syllables, and articulation rate, also have positive effects on engagement, while the average number of pauses had negative effects on engagement. But we address this at the end of the Discussion section.

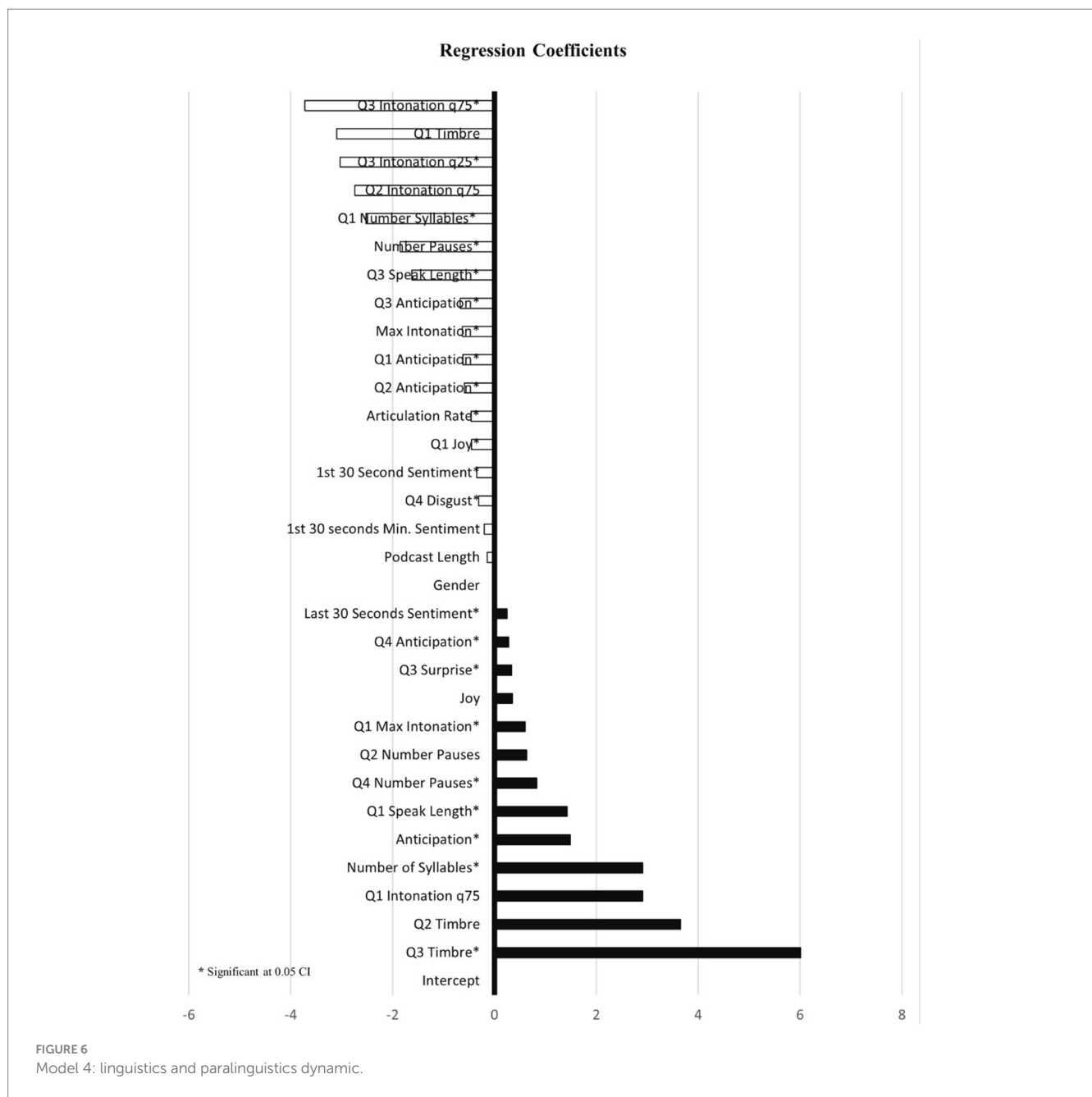
Further explanation of how these dynamic prosodic markers influence engagement is warranted. Research into the interaction between speaker and listener aims to create a shared cognitive representation, which enables the listener during the podcast to make predictions or hypotheses about what was, is, and might be said (Hasson et al., 2012; Huang et al., 2017). In this context, prosodic contours or episodes that elicit curiosity or surprise with varying intonations (Rocklage and Fazio, 2020) or changes in the rate of speech at specific times (Rizzo and Berger, 2023) may have the potential to increase audience engagement. We prefer to compare these prosodic contours to the keys of an organ, which, when activated during certain podcast episodes, allow the audience to verify or falsify their hypotheses or expectations. Similarly, Niebuhr et al. (2016) speak about the orchestration of prosodic features and Berger et al. (2021) speak about emotional volatility. In other terms, not the average prosodic markers matter but the variations being created during different episodes in the speech. Podcasts probably can be conceived as a form of “edutainment,” thus enhancing engagement (Perks et al., 2019; Aksakal, 2015; Oslawski-Lopez and Kordsmeier, 2021).

However, it is important to note that TEDx podcasts represent a specific genre, both distinct from other genres of podcasts such as podcasts on pure entertainment, stories, music, crime, or mental health advice (Tsagkias et al., 2010; Perks and Turner, 2019) and are even more distinct from speeches at political gatherings or religious sermons, where the audience primarily seeks to validate their own worldviews or political preferences. Therefore, the findings of this study are limited to this specific genre of podcasts; future research should explore different podcast genres (Perks and Turner, 2019).

Third, substantiating Conjecture 2, segmenting the podcast presenters by gender added more variance to the prediction model, achieving an adjusted R^2 of 0.50, which accounts for an additional 12% of the explained variance in the female population, and an adjusted R^2 of 0.73, accounting for more than 36% of the explained variance in the male population. The significant increase in explained variance represents another crucial finding of this study. It indicates that both female and male presenters can influence audience engagement by adding prosodic contours or episodes to specific segments of their podcast narratives in distinct ways, thereby enhancing or diminishing engagement. This raises two pertinent questions: (a) what are the reasons for these gender disparities, and (b) has this finding been explored in previous literature?

A detailed analysis of the data from models 6 and 7, which categorize podcasts by gender presenters, reveals significant patterns or trends of how prosodic contours influence audience engagement, either positively or negatively. It is particularly noteworthy that for both female and male presenters, granular elements of surprise and anticipation, in conjunction with distinct paralinguistic features, impact engagement both positively and negatively. This observation aligns with what Baek and Falk (2018) refer to as the “salesperson effect”: podcast listeners seek to acquire new knowledge and share it with their peers. Podcast presenters can enhance the impact of their delivery by strategically segmenting their speech using prosody. They act as strategic or skilled orchestrators of how the content is disseminated.

The literature, exemplified by Perez (2019), underscores linguistic and paralinguistic gender dimorphisms, promoting the separate analysis of data pertaining to males and females. Although acknowledging the complexity of gender dimorphism (Li et al., 2022), several conjectures are proposed regarding the rationale for



researchers undertaking such endeavors. These include hypotheses related to female hormonal influences, such as oxytocin production being more pronounced in females (Schirmer and Kotz, 2006, p. 27), as well as cultural-dependent role expectations (Clarke and Healey, 2022). In this context, various researchers have proposed that females encounter disadvantages when attempting to persuade audiences (McSweeney et al., 2022). However, this observation is not supported consistently (Bapna and Ganco, 2021). Notably, differences in average pitch are often regarded as a primary factor (see Leung et al., 2018). High pitch in female speech is posited to reinforce negative gender biases against female entrepreneurs, whereas lower pitch, irrespective of gender, is associated with greater persuasiveness (Guyer et al., 2023; Clarke and Healey, 2022; Wang et al., 2021). As readers may observe, our machine learning approach enabled us to capture a broader array of paralinguistic features than pitch, which might influence

engagement. Indeed, as Balachandra et al. (2021) note, when females and males communicate in a professional context, both are professionals capable of modulating their prosody to effectively convey their intended messages. We explore this topic further in the following sections.

But first, rather than making general statements about gender difference, the first question that can be asked in this study is: would engagement of the audience for female versus male TEDx podcast presenters be different? The sample in this study consists of 112 female and 76 male podcast presenters. The samples are not equal, therefore we first checked for homoscedasticity in order to examine whether the two samples showed variances in homogeneity. Therefore, a Fisher's *F*-test was performed and we found a *p*-value >0.05, which means the variances of the log of Engagement for Male and Female podcast presenters are homogenous. This allows us to use a two-sample *t*-test

TABLE 8 Model 5 linguistics and paralinguistic and dynamic interactions.

| Variable | Coefficients | p-value |
|---|--------------|---------|
| Intercept | 0.03 | 0.74 |
| Q1 Surprise x Q1 Intonation q75 | 4.76 | 0.00 |
| Number Syllables | 2.63 | 0.00 |
| Podcast Length | 1.55 | 0.07 |
| Anticipation | 0.51 | 0.00 |
| Surprise | 0.34 | 0.02 |
| Q4 Surprise x Q4 Articulation Rate | 0.33 | 0.00 |
| Q1 Negative x Q1 Intonation q25 | 0.29 | 0.00 |
| Q4 Positive x Q4 Articulation Rate | 0.28 | 0.01 |
| Q4 Anticipation x Q4 Articulation Rate | 0.21 | 0.04 |
| Q1 Positive x Q1 Min Intonation | 0.20 | 0.06 |
| Q2 Intonation Variability | 0.19 | 0.06 |
| Articulation Rate | 0.18 | 0.15 |
| Gender | -0.03 | 0.77 |
| Q1 Anger x Articulation Rate | -0.21 | 0.03 |
| Q4 Fear | -0.22 | 0.06 |
| Q3 Surprise x Q3 Articulation Rate | -0.20 | 0.01 |
| Q4 Surprise x Q4 Timbre | -0.29 | 0.01 |
| Q2 Joy x Q2 Timbre | -0.33 | 0.01 |
| Q1 Anticipation | -0.34 | 0.02 |
| Q1 Fear x Q1 Minimum Intonation | -0.37 | 0.02 |
| Q1 Surprise x Q1 Intonation Variability | -0.97 | 0.00 |
| Q1 Number Syllables | -1.41 | 0.00 |
| Q3 Number Syllables | -1.52 | 0.00 |
| Number Pauses | -1.54 | 0.04 |
| Q1 Surprise x Q1 Timbre | -4.05 | 0.00 |

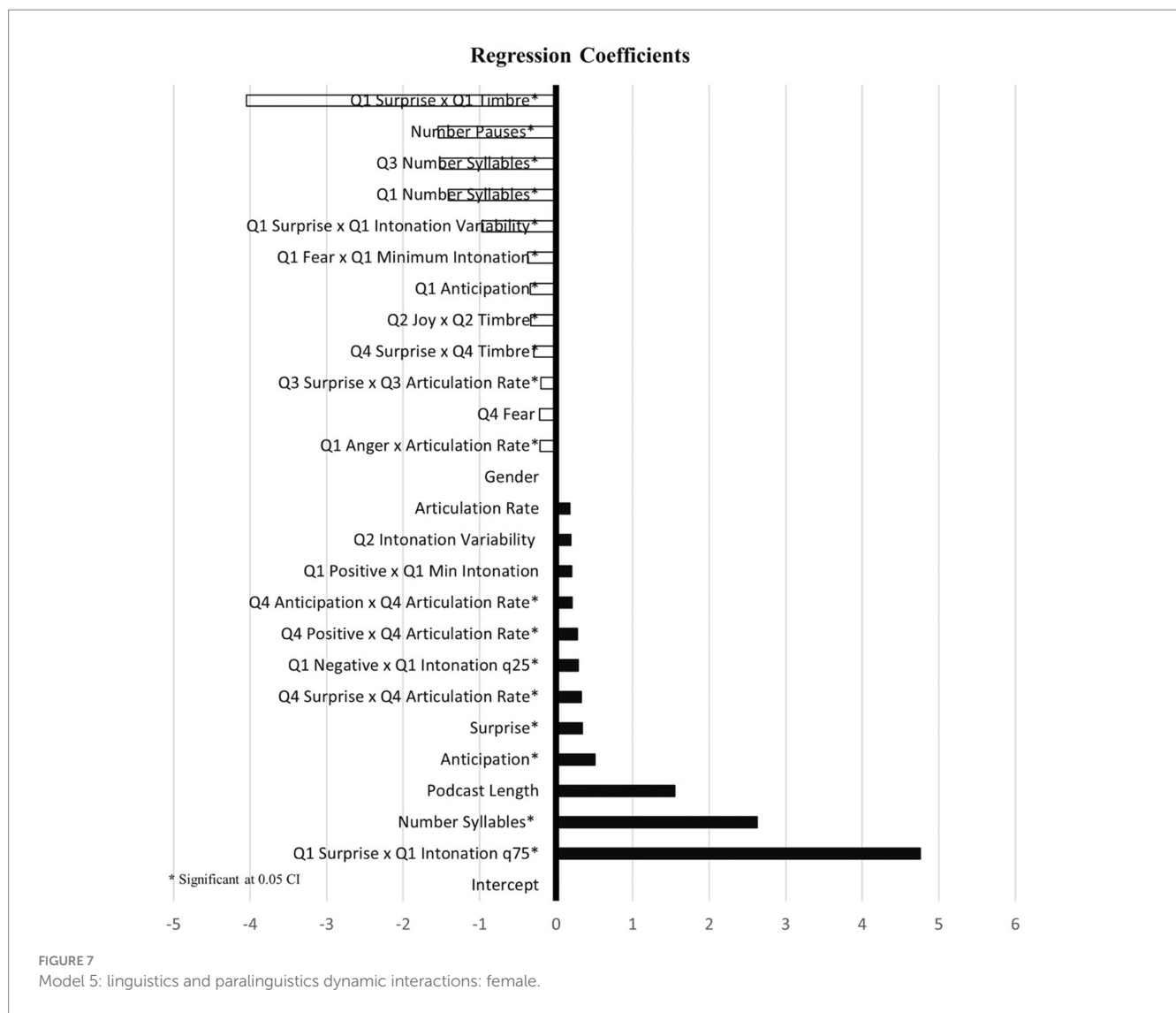
with equal variance version. The 112 female sample was compared for the log of Engagement ($M=4.83$, $SD=1.19$) to the 76 male sample ($M=4.91$, $SD=1.15$) and showed no significant difference, $t(185)=-0.48$, p -value=0.64, despite male speakers showing the higher mean log of engagement. Thus we might conclude that female versus male presenters are not necessarily a disadvantage, at least in this very select sample of world-renowned TEDx speaker experts.

It is important to note that research on the detection of emotions in speech has demonstrated gender disparities in the perception of how prosodic features “leak” the emotions of the speaker (Lausen and Schacht, 2018). The situation, however, is complex. As investigated by Lausen and Schacht (2018), women are often regarded as “emotional experts” and are generally more attuned to categorizing the emotions of speakers in general. But depending on emotional categories, females are especially sensitive to happy, fearful, and neutral emotions and apparently also sensitive to positive emotions expressed by females. Interestingly, both males and females are equally sensitive to anger. This observation leads to a pertinent research question: future studies should not only focus on podcasts in general or the gender of the podcast presenter but also on the composition of the audience. For example, are women more engaged when listening to female speakers? Although our data do not confirm

this, this exploratory study at least provides a foundation for future research.

Other authors, such as Niebuhr et al. (2019), have also examined these gender disparities in speech, but they are adopting an interventionist perspective while focusing on a broader set of paralinguistic features beyond mere pitch. Speech delivery can be trained, and prosodic episodes can be staged. In their study, they introduced both male and female participants to a charisma training program they titled PASCAL: Prosodic Analysis of Speaker Charisma—Assessment and Learning. This program trained participants how to use (average) prosodic features such as pitch, articulation rate, disfluency count, and prosodic phrase durations. They reported that females received faster benefits, but only on specific prosodic features from the prosodic training. Hence, excelling in professional communication is possible.

Although no gender disparities in podcast engagement were observed in our TEDx data, our study suggests that speech trainers or coaches working with both male and female speakers could improve training by focusing more on the use of dynamic or granular prosodic contours rather than on average paralinguistic markers. These techniques involve two primary strategies: (a) envisioning a well-defined plot or script to provide the speech structure, and (b) emphasizing specific points during a speech or creating more engaging



episodes that incorporate both anticipation and surprise, thereby utilizing prosodic contours as triggers for hypothesis generation and testing by the audience.

Two final thoughts. As observant readers may have noticed, the length of the podcasts in most models presented in this study influenced audience engagement. Given that these TEDx podcasts are consumed in diverse settings—such as during leisure, travel, or while performing chores—it is crucial that the content or segments of the podcast are appropriately timed to align with the specific durations and motivations of the audience, whether these are brief or extended. Consequently, further research is warranted to investigate the impact of podcast length on audience engagement in the specific contexts in which they are consumed.

Furthermore, while focusing on affective linguistic features, we employed a specific sentiment scoring tool, namely the NRC Emotion Lexicon (Mohammad and Turney, 2013), as it is frequently employed in the machine learning literature (Xue et al., 2020). Other language tools could have been used, such as LIWC, which includes both affective and general linguistic features. Alternatively, for affective language features, we could have used tools based on the circumplex model of affect by Russell (1980), as employed by Lekkas

et al. (2022). Future research may utilize these text analysis tools to further explore the research inquiries addressed in this study.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform’s terms of use and all relevant institutional/national regulations.”

TABLE 9 Model 6 linguistics and paralinguistics dynamic interactions: female.

| Variable | Coefficients | p-value |
|--|--------------|---------|
| Intercept | -0.03 | 0.79 |
| Q3 Anger x Q3 Timbre | 1.87 | 0.04 |
| Q3 Disgust x Q3 Intonation q25 | 1.55 | 0.04 |
| Q3 Negative x Q3 Articulation Rate | 0.68 | 0.00 |
| Q1 Anticipation x Q1 Intonation Var. | 0.42 | 0.00 |
| Q4 Surprise x Q4 Articulation Rate | 0.37 | 0.04 |
| Q4 Anticipation | 0.34 | 0.01 |
| Q1 Joy x Q1 Timbre | 0.27 | 0.06 |
| Q4 Anticipation x Q4 Articulation Rate | 0.25 | 0.12 |
| Q3 Surprise | 0.25 | 0.06 |
| Articulation Rate | 0.22 | 0.11 |
| Q2 Joy x Q2 Timbre | -0.22 | 0.06 |
| Q4 Positive x Q4 Max. Intonation | -0.29 | 0.04 |
| Q3 Surprise x Q3 Articulation Rate | -0.25 | 0.01 |
| Podcast length | -0.36 | 0.01 |
| Q4 Surprise x Q4 Intonation q25 | -0.37 | 0.03 |
| Q4 Surprise x Q4 Min. Intonation | -0.37 | 0.02 |
| Q1 Anger x Q1 Timbre | -0.38 | 0.01 |
| Q3 Disgust x Q3 Timbre | -1.34 | 0.01 |
| Q3 Anger x Q3 Intonation q25 | -1.49 | 0.11 |

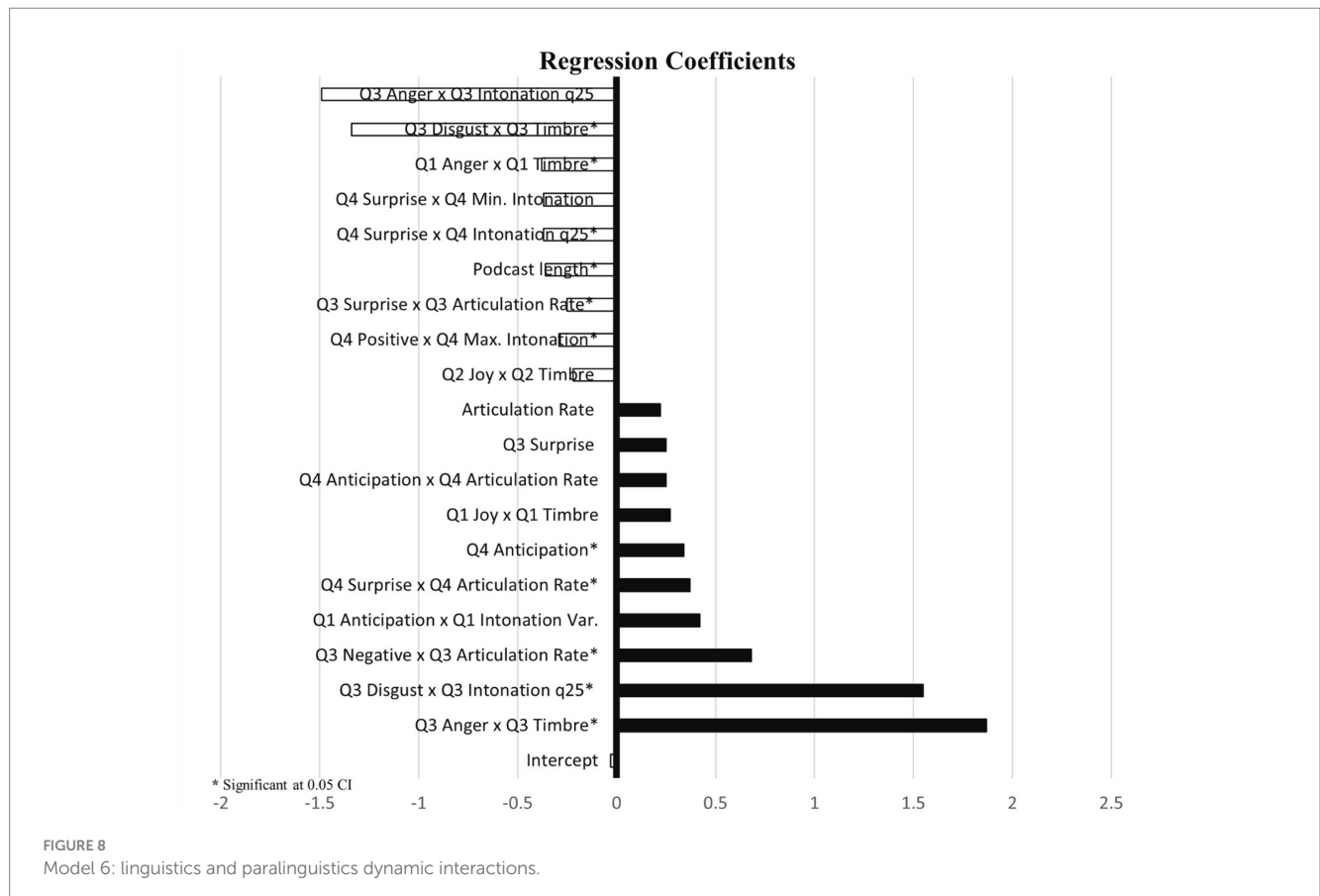


TABLE 10 Model 7 linguistics and paralinguistics dynamic interactions: male.

| Variable | Coefficients | p-value |
|--|--------------|---------|
| Intercept | -0.00 | 0.99 |
| Q2 Negative x Q2 Articulation Rate | 1.49 | 0.00 |
| Number Pauses | 1.21 | 0.00 |
| Q3 Anger x Q3 Articulation Rate | 1.11 | 0.00 |
| Number Syllables | 0.99 | 0.07 |
| 1 st 30 Second Sentiment Score Var. | 0.67 | 0.00 |
| Q3 Surprise x Q3 Max Intonation | 0.64 | 0.00 |
| Q3 Anticipation x Q3 Articulation Rate | 0.64 | 0.00 |
| Q3 Anticipation x Q3 Timbre | 0.58 | 0.00 |
| Positive vs. Negative Ratio | 0.54 | 0.00 |
| Q2 Sadness x Q2 Articulation Rate | 0.53 | 0.03 |
| Q4 Positive x Q4 Articulation Rate | 0.42 | 0.00 |
| Q2 Fear x Q2 Max Intonation | 0.29 | 0.08 |
| Q4 Negative x Q4 Max Intonation | 0.27 | 0.06 |
| Q4 Trust x Q4 Intonation q75 | -0.26 | 0.08 |
| Q1 Surprise x Q1 Min Intonation | -0.26 | 0.04 |
| Q1 Sadness x Q1 Max Intonation | -0.26 | 0.10 |
| Q1 Surprise x Q1 Articulation Rate | -0.27 | 0.11 |
| Q1 Surprise x Q1 Articulation Rate | -0.30 | 0.06 |
| Q1 Trust x Q1 Timbre | -0.35 | 0.01 |
| Q3 Joy x Q3 Articulation Rate | -0.45 | 0.00 |
| Q1 Surprise x Q1 Intonation Variability | -0.48 | 0.00 |
| Q3 Surprise x Q3 Timbre | -0.52 | 0.01 |
| Q3 Surprise x Q3 Articulation Rate | -0.55 | 0.01 |
| Q2 Disgust x Q2 Articulation Rate | -0.55 | 0.06 |
| Q4 Sadness x Q4 Max Intonation | -0.86 | 0.00 |
| Word Count | -0.92 | 0.01 |
| Q3 Surprise x Q3 Articulation Rate | -0.93 | 0.00 |
| Podcast Length | -1.00 | 0.05 |
| Q2 Disgust x Q2 Articulation Rate | -3.43 | 0.00 |

Author contributions

AS: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing, Visualization. WV: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

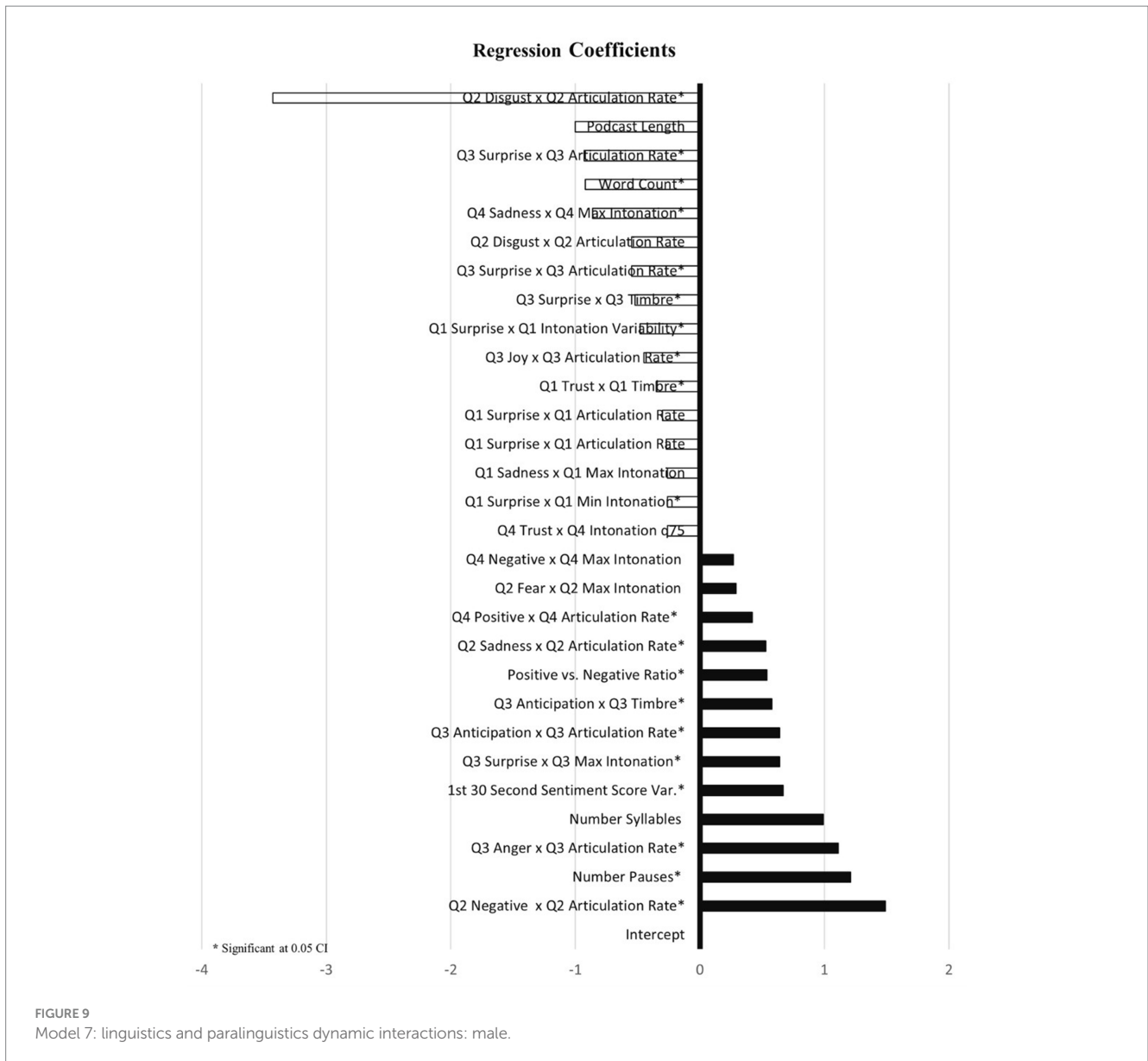


TABLE 11 Predictive models with various parameters.

| Model | Residual standard error | Degree of freedom | Multiple R squared | Adjusted R squared | F-statistic | p-value |
|--|-------------------------|-------------------|--------------------|--------------------|-------------|---------|
| Model 1 Linguistic- Average | 1.499 | 173 | 0.101 | 0.059 | 2.429 | 0.02 |
| Model 2 Linguistic and Paralinguistic Average | 1.489 | 168 | 0.139 | 0.073 | 2.090 | 0.02 |
| Model 3 Linguistic Dynamic | 1.365 | 153 | 0.341 | 0.221 | 2.829 | 0.00 |
| Model 4 Linguistic and Paralinguistic Dynamic | 1.327 | 150 | 0.389 | 0.263 | 3.078 | 0.00 |
| Model 5 Linguistics and Paralinguistic and Dynamic Interactions | 1.220 | 156 | 0.463 | 0.377 | 5.378 | 0.00 |
| Model 6 Linguistics and Paralinguistics Dynamic Interactions: Female | 1.094 | 89 | 0.589 | 0.501 | 6.706 | 0.00 |
| Model 7 Linguistics and Paralinguistics Dynamic Interactions: Male | 0.809 | 43 | 0.838 | 0.729 | 7.669 | 0.00 |

References

- Aksakal, N. (2015). Theoretical view to the approach of the edutainment. *Procedia-Social Beh. Sci.* 186, 1232–1239. doi: 10.1016/j.sbspro.2015.04.081
- Anderson, R. C., Klostfald, C. A., Mayew, W. J., and Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS One* 9:e97506. doi: 10.1371/journal.pone.0097506
- Aung, T., and Puts, D. (2020). Voice pitch: a window into the communication of social power. *Cur. Op. Psy.* 33, 154–161. doi: 10.1016/j.copsyc.2019.07.028
- Backlinko. (2024). Available at: <https://backlinko.com/podcast-stats> (Accessed September, 2024)
- Baek, E. C., and Falk, E. B. (2018). Persuasion and influence: what makes a successful persuader? *Curr. Op. Psy.* 24, 53–57. doi: 10.1016/j.copsyc.2018.05.004
- Balachandra, L., Briggs, T., Eddleston, K., and Brush, C. (2019). Don't pitch like a girl!: how gender stereotypes influence investor decisions. *Entrepr. Th. Prac.* 15, 116–137. doi: 10.1177/1042258717728028
- Balachandra, L., Fischer, K., and Brush, C. (2021). Do (women's) words matter? The influence of gendered language in entrepreneurial pitching. *J. Bus. Ventur. Insights* 15:e00224. doi: 10.1016/j.jbvi.2021.e00224
- Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Comm.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016
- Bapna, S., and Ganco, M. (2021). Gender gaps in equity crowdfunding: evidence from a randomized field experiment. *Manag. Sci.* 67, 2679–2710. doi: 10.1287/mnsc.2020.3644
- Berger, J. (2023). *Magic words: Bentang Pustaka*. New York, USA: Harper Collins Publishing
- Berger, J., Kim, Y. D., and Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *J. Constr. Res.* 48, 235–250. doi: 10.1093/jcr/ucab010
- Berger, J., Packard, G., Boghrati, R., Hsu, M., Humphreys, A., Luangrath, A., et al. (2022). Wisdom from words: marketing insights from text. *Mark. Lett.* 33, 365–377. doi: 10.1007/s11002-022-09635-6
- Berry, M., and Brown, S. (2019). Acting in action: prosodic analysis of character portrayal during acting. *J. Exp. Psy. Gen* 148, 1407–1425. doi: 10.1037/xge0000624
- Brinson, N. H., and Lemon, L. L. (2023). Investigating the effects of host trust, credibility, and authenticity in podcast advertising. *J. Mark. Commun.* 29, 558–576. doi: 10.1080/13527266.2022.2054017
- Cascio Rizzo, G. L., Berger, J., De Angelis, M., and Pozharliev, R. (2023). How sensory language shapes Influencer's impact. *J. Constr. Res.* 50, 810–825. doi: 10.1093/jcr/ucad017
- Chang, H. H., Mukherjee, A., and Chattopadhyay, A. (2023). More voices persuade: the attentional benefits of voice numerosity. *J. Mark. Res.* 60, 687–706. doi: 10.1177/00222437221134115
- Chan-Olmsted, S., and Wang, R. (2022). Understanding podcast users: consumption motives and behaviors. *New Media Soc.* 24, 684–704. doi: 10.1177/1461444820963776
- Clarke, J., and Healey, M. P. (2022). "Giving voice to persuasion: embodiment, the voice and cultural entrepreneurship" in *Advances in cultural entrepreneurship*, vol. 80 (Leeds, England: Emerald Publishing Limited), 37–56.
- Cosimini, M. J., Cho, D., Liley, F., and Espinoza, J. (2017). Podcasting in medical education: how long should an educational podcast be? *J. Grad. Med. Ed.* 9, 388–389. doi: 10.4300/JGME-D-17-00015.1
- Daly, N., and Warren, P. (2001). Pitching it differently in New Zealand English: speaker sex and intonation patterns. *J. Socioling.* 5, 85–96. doi: 10.1111/1467-9481.00139
- Dowling, D. O., and Miller, K. J. (2019). Immersive audio storytelling: podcasting and serial documentary in the digital publishing industry. *J. Rad. Aud. Med.* 26, 167–184. doi: 10.1080/19376529.2018.1509218
- Gangamohan, P., Kadiri, S. R., and Yegnanarayana, B. (2016). Analysis of emotional speech—a review. *Toward Robotic Socially Believable Behaving Systems-Volume I: Mod. Em.* 205–238. doi: 10.1007/978-3-319-31056-5_11
- Grice, H. P. (1957). Meaning. *Philos. Rev.* 66, 377–388. doi: 10.2307/2182440
- Guyer, J. J., Briñol, P., Vaughan-Johnston, T. I., Fabrigar, L. R., Moreno, L., Paredes, B., et al. (2023). Pitch as a recipient, channel, and context factor affecting thought reliance and persuasion. *Personal. Soc. Psychol. Bull.* doi: 10.1177/01461672231197547
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Tr. Cog. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007
- Henton, C. (1995). Pitch dynamism in female and male speech. *Lang. Comput.* 15, 43–61. doi: 10.1016/0271-5309(94)00011-Z
- Hollebeek, L. D. (2019). Developing business customer engagement through social media engagement-platforms: an integrative SD logic/RBV-informed model. *Ind. Mark. Manag.* 81, 89–98. doi: 10.1016/j.indmarman.2017.11.016
- Hovy, D., Melumad, S., and Inman, J. J. (2021). Wordify: a tool for discovering and differentiating consumer vocabularies. *J. Constr. Res.* 48, 394–414. doi: 10.1093/jcr/ucab018
- Huang, K., Yeomans, M., Brooks, A. W., Minson, J., and Gino, F. (2017). It doesn't hurt to ask: question-asking increases liking. *J. Pers. Soc. Psychol.* 113, 430–452. doi: 10.1037/pspi0000097
- Insider Intelligence. (2024). Available at: <https://www.businessinsider.com/intelligence/the-podcast-audience>.
- Jeon, Y. A., Son, H., Chung, A. D., and Drumwright, M. E. (2019). Temporal certainty and skippable in-stream commercials: effects of ad length, timer, and skip-ad button on irritation and skipping behavior. *J. Int. Mark.* 47, 144–158. doi: 10.1016/j.intmar.2019.02.005
- Ko, S. J., Judd, C. M., and Blair, I. V. (2006). What the voice reveals: within- and between-category stereotyping on the basis of voice. *Personal. Soc. Psychol. Bull.* 32, 806–819. doi: 10.1177/0146167206286627
- Lausen, A., and Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Front. Psychol.* 9:359771. doi: 10.3389/fpsyg.2018.00882
- Lekkas, D., Gyorda, J. A., Price, G. D., Wortzman, Z., and Jacobson, N. C. (2022). Using the COVID-19 pandemic to assess the influence of news affect on online mental health-related search behavior across the United States: integrated sentiment analysis and the circumplex model of affect. *J. Med. Internet Res.* 24:2731. doi: 10.2196/32731
- Leung, Y., Oates, J., and Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: a systematic review and meta-analysis. *J. Spe. Lang. Hear. Res.* 61, 266–297. doi: 10.1044/2017_JSLHR-S-17-0067
- Li, Y., Sui, S., and Wu, S. (2022). The effect of gender fit on crowdfunding success. *J. Bus. Ventur. Insights* 18:e00333. doi: 10.1016/j.jbvi.2022.e00333
- Lindgren, M. (2023). Intimacy and emotions in podcast journalism: a study of award-winning Australian and British podcasts. *Journal. Pract.* 17, 704–719. doi: 10.1080/17512786.2021.1943497
- Liu, S., Si, G., and Gao, B. (2022). Which voice are you satisfied with? Understanding the physician-patient voice interactions on online health platforms. *Decis. Support. Syst.* 157:113754. doi: 10.1016/j.dss.2022.113754
- Luangrath, A. W., Xu, Y., and Wang, T. (2023). Paralinguistic classifier (PARA): an algorithm for automatic coding of paralinguistic nonverbal parts of speech in text. *J. Mark. Res.* 60, 388–408. doi: 10.1177/00222437221116058
- McSweeney, J. J., McSweeney, K. T., Webb, J. W., and Devers, C. E. (2022). The right touch of pitch assertiveness: examining entrepreneurs' gender and project category fit in crowdfunding. *J. Bus. Ventur.* 37:106223. doi: 10.1016/j.jbusvent.2022.106223
- Mohammad, S. M., and Turney, P. D. (2013). NRC emotion lexicon. *Natl. Res. Counc. Can.* 2:234. doi: 10.1109/MIS.2008.57
- Niebuhr, O., Tegtmeier, S., and Schweisfurth, T. (2019). Female speakers benefit more than male speakers from prosodic charisma training—a before-after analysis of 12-weeks and 4-h courses. *Front. Comm.* 4:427223. doi: 10.3389/fcomm.2019.00012
- Niebuhr, O., Voße, J., and Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve jobs tone of voice. *Comput. Hum. Behav.* 4, 366–382. doi: 10.1016/j.chb.2016.06.059
- Novák-Tóth, E., Niebuhr, O., and Chen, A. (2017). "A gender bias in the acoustic-melodic features of charismatic speech?" in *Proc. Int. Conf. Spok. Lang. Processing*, vol. 12 (International Speech Communication Association), 2248–2252.
- O'Keefe, D. J., and Jensen, J. D. (2007). The relative persuasiveness of gain-framed loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. *J. Health Commun.* 12, 623–644. doi: 10.1080/10810730701615198
- Oslawski-Lopez, J., and Kordsmeier, G. (2021). "Being able to listen makes me feel more engaged": best practices for using podcasts as readings. *Teach. Sociol.* 49, 335–347. doi: 10.1177/0092055X211017197
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv Preprint*.
- Perez, C. C. (2019). *Invisible women: Data bias in a world designed for men*. New York, USA: Abrams.
- Perks, L. G., and Turner, J. S. (2019). Podcasts and productivity: a qualitative uses and gratifications study. *Mass Commun. Soc.* 22, 96–116. doi: 10.1080/15205436.2018.1490434
- Perks, L. G., Turner, J. S., and Tollison, A. C. (2019). Podcast uses and gratifications scale development. *J. Broadcast. Electron. Media.* 63, 617–634. doi: 10.1080/08838151.2019.1688817
- Pernet, C. R., and Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Front. Psychol.* 3:11939. doi: 10.3389/fpsyg.2012.00023
- Petty, R. E., and Cacioppo, J. T. (2012). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, USA: Springer Science and Business Media.
- Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190. doi: 10.1017/S0140525X04000056

- Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., and Reby, D. (2018). Voice pitch modulation in human mate choice. *Proc. R. Soc. B* 285:20181634. doi: 10.1098/rspb.2018.1634
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *J. Acoust. Soc. Am.* 123, 1104–1113. doi: 10.1121/1.2821762
- Rizzo, C., and Berger, J. A. (2023). The power of speaking slower. Available at: <https://ssrn.com/abstract=4580994> (Accessed September 23, 2023).
- Rocklage, M. D., and Fazio, R. H. (2020). The enhancing versus backfiring effects of positive emotion in consumer reviews. *J. Mark. Res.* 57, 332–352. doi: 10.1177/0022243719892594
- Rocklage, M. D., Rucker, D. D., and Nordgren, L. F. (2018). The evaluative lexicon 2.0: the measurement of emotionality, extremity, and valence in language. *Behav. Res. Methods* 50, 1327–1344. doi: 10.3758/s13428-017-0975-6
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Schirmer, A., and Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. *Tr. Cog. Sci.* 21, 216–228. doi: 10.1016/j.tics.2017.01.001
- Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Tr. Cog. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009
- Schwartz, R., and Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics. *PLoS One* 7:e47279. doi: 10.1371/journal.pone.0047279
- Sotério, C., and Linhares Queiroz, S. (2023). Improving writing skills through scripting a science podcast for non-expert audiences. *J. Coll. Sci. Teach.* 52, 30–37. doi: 10.1080/0047231X.2023.12315864
- Tausczik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* 29, 24–54. doi: 10.1177/0261927X09351676
- Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., and Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Ev. H. Beh.* 33, 210–216. doi: 10.1016/j.evolhumbehav.2011.09.004
- Truninger, M., Ruderman, M. N., Clerkin, C., Fernandez, K. C., and Cancro, D. (2021). Sounds like a leader: an ascription–actuality approach to examining leader emergence and effectiveness. *Leadersh. Q.* 32, 101420. doi: 10.1016/j.leaqua.2020.101420
- Tsagkias, M., Larson, M., and De Rijke, M. (2010). Predicting podcast preference: an analysis framework and its application. *J. Am. Soc. Inf. Technol.* 61, 374–391. doi: 10.1002/asi.21259
- Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054
- Van Krieken, K., Hoeken, H., and Sanders, J. (2015). From reader to mediated witness: the engaging effects of journalistic crime narratives. *J. Mass Comm. Quart.* 92, 580–596. doi: 10.1177/1077699015586546
- Van Zant, A. B., and Berger, J. (2020). How the voice persuades. *J. Pers. Soc. Psychol.* 118, 661–682. doi: 10.1037/pspi0000193
- Wang, X., Lu, S., Li, X. I., Khamitov, M., and Bendle, N. (2021). Audio mining: the role of vocal tone in persuasion. *J. Constr. Res.* 48, 189–211. doi: 10.1093/jcr/ucab012
- Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., et al. (2020). Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. *J. Med. Internet Res.* 22:e20550. doi: 10.2196/20550