

High-Energy Ad Content: A Large-Scale Investigation of TV Commercials

Journal of Marketing Research
1-20
© American Marketing Association 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00222437211067802
journals.sagepub.com/home/mrj



Joonhyuk Yang , Yingkang Xie, Lakshman Krishnamurthi,
and Purushottam Papatla

Abstract

A trend reported by both academics and practitioners is that advertising on TV has become increasingly energetic. This study investigates the association between the energy level in ad content and consumers' ad-tuning tendency. Using a data set of over 27,000 TV commercials delivered to U.S. homes between 2015 and 2018, the authors first present a framework to algorithmically measure the energy level in ad content from the video of ads. This algorithm-based measure is then compared with human-perceived energy levels showing that the measure is related to the level of arousal stimulated by ad content. By relating the energy levels in ad content with the ad-tuning tendency using two empirical procedures, the authors document the following: overall, more energetic commercials are more likely to be tuned in or less likely to be avoided by viewers. The positive association between energy levels in ad content and ad tuning is statistically significant after controlling for placement and other aspects of commercials. However, the association varies across product categories and program genres. The main implication of this study is that advertisers should pay attention to components of ad content other than loudness, which has been regulated by law.

Keywords

advertising, ad tuning, ad avoidance, multimedia data, audio information retrieval

Online supplement: <https://doi.org/10.1177/00222437211067802>

Although consumers' responses to particular choices of ad content have long been studied (e.g., Olney, Holbrook, and Batra 1991; Preston 1982), only recently have researchers started to use computational methods with large-scale, unstructured, nonlaboratory data.¹ With increasing progress in data and methodology, marketers now have the opportunity to better understand the influence of ad content (which is inherently unstructured) on consumer behaviors, leading to better design of ads.

In this study, we focus on one aspect of ad content that has captured both researchers' and practitioners' attention, namely, high-energy stimuli in advertising. The energy of an ad is related to how stimulating it is. While there is no standard definition of "energy" in advertising, Puccinelli, Wilcox, and Grewal (2015) state, "High-energy commercials are television ads that are active, exciting and arousing for the viewer to experience" (p. 1). High-energy commercials are reported to have become prevalent in the TV

advertising market. For instance, Puccinelli, Wilcox, and Grewal (2015) found that more than 80% of commercials on Hulu were rated as relatively energetic. Our investigation of Super Bowl ads as well as TV commercials on the major broadcast networks also confirms that the energy level of ad content has increased.

Motivated by these observations, we examine whether the use of high energy in TV ads affects viewer behavior in terms of how long ads are tuned in. Studies in marketing have demonstrated that the inclusion of highly arousing stimuli increases viewers' attention to ads (e.g., Belanche, Flavián, and Pérez-Rueda 2017). Other research, however, points to the opposite, namely

¹ In a detailed analysis of unstructured data in marketing, Balducci and Marinova (2018) describe unstructured data as nonnumeric, with no predefined representation of the construct of interest, as multifaceted, and as providing concurrent information.

Joonhyuk Yang is Assistant Professor of Marketing, Mendoza College of Business, University of Notre Dame, USA (email: joonhyuk.yang@nd.edu). Yingkang Xie is a Ph.D. candidate in Marketing, Kellogg School of Management, Northwestern University, USA (email: yingkang.xie@kellogg.northwestern.edu). Lakshman Krishnamurthi is A. Montgomery Ward Professor of Marketing, Kellogg School of Management, Northwestern University, USA (email: laksh@kellogg.northwestern.edu). Purushottam Papatla is Northwestern Mutual Data Science Institute Professor of Marketing and Co-Director, Northwestern Mutual Data Science Institute, Lubar School of Business, University of Wisconsin-Milwaukee, USA (email: papatla@uwm.edu).

that high energy can yield a negative outcome under certain conditions (e.g., Puccinelli, Wilcox, and Grewal 2015). Given these different possibilities, we empirically evaluate whether and how higher energy in TV ads, overall, is associated with viewers' ad-tuning rates.

A key differentiating feature of our research is that, unlike previous studies conducted in lab settings, we utilize a large-scale data set of actual TV commercials tuned in at people's homes.² Specifically, our data set contains detailed information on the insertions of over 27,000 ads on the five major broadcast networks in the United States (ABC, CBS, FOX, NBC, and The CW) during the three-year period between September 2015 and August 2018. In all, our data set includes over a million insertions of the ads with information on the network, program, date, and time of each airing, as well as the video of each ad. For each insertion of an ad, we have data on the percentage of viewers who stayed tuned in for at least 25%, 50%, and 75% of the ad's duration. We use these measures to relate the energy in ad content to the tuning rate of ads. These measures are relevant to advertisers because so-called call-to-action or selling arguments typically come at the end of ads, following attention-grabbing content in the earlier parts (Teixeira, Picard, and Kaliouby 2014).

One empirical challenge is to measure each ad creative's energy level. Employing human coders would be both costly and infeasible given the large number of ads we aim to investigate, especially when many coders would be needed to rate each ad to ensure measurement reliability. Human ratings are also susceptible to individual factors and contextual factors such as the specific location, time of measurement, and preexisting mood. We overcome this challenge using an algorithm-based approach. Specifically, we operationalize energy in ad content by adapting a widely used proprietary approach from the music streaming service Spotify. Spotify defines energy as "a perceptual measure of intensity and powerful activity released throughout the track. Typical energetic tracks feel fast, loud, and noisy" (Jehan and Montecchio 2016). Notably, Spotify's energy measure has been formally investigated in computer science (Ferraro, Bogdanov, and Serra 2019; Ng and Mehrotra 2020), sociology (Askin and Mauskopf 2017), and marketing (Boughanmi, Ansari, and Kohli 2021).

Nonetheless, the scalability and efficiency gains from using an algorithm-based approach also come with a cost. First and foremost, we need a better understanding of the relationship between the Spotify energy metric (an objective measure based on low-level auditory features) and viewers' perceived affect in our context. To this end, we conducted a separate analysis using data collected via Amazon Mechanical Turk (MTurk). Borrowing the circumplex model of affect from Russell (1980), we find that our algorithm-generated energy

measure is positively correlated with the two dimensions of affect, namely arousal and valence.³ Citing Mehrabian and Russell (1974), Di Muro and Murray (2012, p. 574) define arousal as "the subjective experience of energy mobilization, which can be conceptualized as an affective dimension ranging from sleepy to frantic excitement." Valence, on the other hand, is "the extent to which an affective state is positive or negative." We find that the correlation is much stronger with arousal (.46–.66) than with valence (.02–.21).

Second, Spotify's algorithm was originally designed for music, so it can only measure the energy level of the auditory component of ads. Obviously, TV commercials consist of not only audio but video as well. Thus, if the energy level in an ad's audio differs substantially from that in its video or its overall energy level, our objective measure may not represent the actual consumer experience. However, results from the MTurk study confirm that perceived arousal from ad audio is well correlated with that from ad video (correlations ranging between .53 and .57) and, more importantly, also highly correlated with the overall perceived arousal from both ad audio and video (correlations ranging between .68 and .86).

Third, because of the proprietary nature of Spotify's measure of energy, we develop our own approach to predict the metric for the ads in our data set. We use a publicly available music data set of over 13,000 songs from the Free Music Archive (FMA), which includes the primitives of audio characteristics characterized by low-level attributes like frequency and spectral flux as well as the Spotify measure of energy for each song. We employ machine learning methods to relate the low-level auditory attributes of each song to the Spotify energy measure for that song. Next, we extract the audio from the videos of each of the ads in our data set and measure the primitive audio characteristics of the extracts. Finally, we use the relationship between energy and the primitive audio characteristics derived from the FMA data set to estimate the audio energy in each ad creative.

We then empirically investigate the relationship between the tuning rates of ad insertions and their energy levels via a two-stage regression. The first stage includes fixed effects of the ads and insertion variables that include the network, program, date, time, and position of the insertion within the program. Controlling for these aspects of ad insertions is important because tuning rates are likely to be context dependent. The second stage takes the estimated fixed effects of the ads as proxies for their quality and regresses them on the derived Spotify energy measure and other creative characteristics of the ads, such as their mood and whether they include a song. The results indicate that, on average, more energetic commercials are more likely to be tuned in for longer or less likely to be avoided by viewers, supporting the positive association.

² We distinguish the term "tuning" from "viewing," "watching," and "attention." Tuning in when an ad is played does not necessarily mean viewing, watching, or paying attention to the ad (e.g., see McGranaghan, Liaukonyte, and Wilbur 2021 on how the rates vary for these behaviors). In our data, we observe only tuning.

³ The arousal–valence framework from Russell (1980) provides a mapping of emotions onto a two-dimensional space, where the two axes are defined as arousal and valence. The model has been widely adopted by researchers to measure affective state (for more discussion, see e.g., Fong, Kumar, and Sudhir [2021]).

However, the associations between energy levels in ad content and the extent of ad tuning vary in both magnitude and direction across product categories and program genres.

To further assess the reliability of this evidence supporting the positive association of energy with tuning rates, we conduct a novel empirical investigation that draws its inspiration from ideal *exogenous* variation across or within ads in their energy levels, while holding other aspects constant. Specifically, we take advantage of the fact that people's preferences for energetic audio exhibit temporal variations over a day (Park et al. 2019). For instance, people generally prefer more energetic audio during the day and less energetic audio at night. Thus, showing an ad, which has a fixed energy level, at different times of the 24-hour day generates variation in the distance from its energy level to the level people generally prefer (i.e., the baseline) at those different times. We utilize an auxiliary data set of Spotify listening sessions, the Music Streaming Sessions Dataset (MSSD; Brost, Mehrotra, and Jehan 2019), to quantify this baseline. We then leverage the time-varying baseline levels to quantify the relative distance from the baseline to each ad's energy level, which generates within-ad variation in audio energy relative to the baseline. We relate the within-ad variation to tuning rates. This approach allows us to jointly estimate the ad fixed effects and energy effects in a single step. We find that higher energy is again associated with longer tuning in or lower ad avoidance, which provides additional support for the positive association.

Our research makes two contributions to the literature. First, we use large-scale data to empirically demonstrate that high-energy ad content is associated with higher tuning rate of ads. On this topic, previous research, heavily reliant on lab studies, finds a positive relationship between arousal and viewing time of ads (Belanche, Flavián, and Pérez-Rueda 2017; Olney, Holbrook, and Batra 1991), while other studies find certain conditions in which increased arousal does not lead to an increase either in attention or in viewing time.⁴

To the best of our knowledge, our work is one of a few studies that connect the tendency of ad avoidance to a particular aspect of ad content using large-scale, nonlaboratory data; others include Wilbur (2008) and McGranaghan, Liaukonyte, and Wilbur (2021). More broadly, our study also adds to the rich marketing literature on ad avoidance in the context of television advertising (e.g., Danaher 1995; Deng and Mela 2018; Siddarth and Chattopadhyay 1998; Teixeira, Wedel, and Pieters 2010; Tuchman, Nair, and Gardete 2018; Van Meurs 1998; Wilbur 2016) by adding the role of a new dimension (i.e., energy in ad content) to the literature.

Second, we introduce a novel approach to measure the energy level of ad content, borrowing audio information retrieval methods from computer science and adapting the

methods for a marketing context, thus adding to the literature on the use of unstructured, multimedia data in marketing research. The volume of marketing research that leverages unstructured multimedia data is ever increasing.⁵ We believe that our study showcases the importance of parsimonious feature engineering of ad content in research relating it to consumer behavior. Algorithm-based approaches to ad content can easily end up with a complex, high-dimensional feature set. Given all the techniques available, one could extract thousands of auditory and visual elements of ads but find that very few are easily interpretable and capable of explaining meaningful variation in consumer behavior (e.g., McGranaghan, Liaukonyte, and Wilbur 2021). In some sense, our approach goes in the opposite direction by carefully choosing a single human-interpretable variable, namely energy in ad content, examining the construct behind it, and investigating its relevance to consumer behavior. The work of Fong, Kumar, and Sudhir (2021) is also in a similar vein as they begin with a theory (i.e., Russell's arousal–valence model of affect) to guide their feature engineering.

Data

We use data from multiple sources as summarized in Table 1. The primary data set for our research is provided by iSpot.tv (www.ispot.tv), which tracked over 9 million internet-connected televisions in the United States during the studied period of the three television seasons from September 2015 through August 2018.⁶ The company relies on automatic content recognition (ACR) technology provided by other companies to detect whether an ad is being played on the screen and, if it is, which brand is being advertised and what creative is being shown. The ACR software installed on each smart TV tracks the ads either played live or time-shifted and played from a DVR-type recording within seven days of the original broadcast. Staff of iSpot.tv reported to us that about 6% of the viewership is time-shifted in our data. Because information is collected once every second from each tracked television, iSpot.tv is able to detect the exact instant at which the ad is interrupted. Specifically, the exact times of actions like fast-forwarding, switching to a different channel, pulling up the program menu, or turning the TV off while an ad is playing are recorded. The identity of the household member who takes those actions is unknown, as we do not have access to household-level data.

The data set we investigate in this research includes all nationally telecast ads available from iSpot.tv during the three

⁴ For instance, Puccinelli, Wilcox, and Grewal (2015) find that energetic advertising will lead to less watching when the context of the program is one of low arousal. Gorn, Pham, and Sin (2001) did not consider tuning rates but find that increased arousal may not always increase ad evaluation.

⁵ This research includes studies on text (e.g., Liu, Singh, and Srinivasan 2016; Netzer et al. 2012), images (e.g., Jalali and Papatla 2016; Liu, Dzyabura, and Mizik 2020; Xiao and Ding 2014), voice (e.g., Marinova, Singh, and Singh 2018; Xiao, Kim, and Ding 2013), and videos (e.g., Li, Shi, and Wang 2019; Lu, Xiao, and Ding 2016), among others.

⁶ Documentation from iSpot.tv describes that the population of tracked televisions is adjusted to ensure that the number of monitored televisions in each designated market area (DMA) and zip code of the country reflects the proportion of all TVs in the country present in that DMA and zip code.

Table 1. Data Description.

Source	Data Set	Observations	Observation Period
iSpot.tv	Ad insertion, videos, and metadata	Ad insertion information (e.g., network, program, date, time), videos and metadata (e.g., duration, and other attributes)	September 2015–August 2018
iSpot.tv	Ad avoidance	Proprietary measure of the likelihood of ad interruption at each ad insertion	September 2015–August 2018
adland.tv	Super Bowl commercials	Videos of Super Bowl commercials	Super Bowl III (1969)–Super Bowl LIV (2020)
Free Music Archive (FMA)	FMA database	Information on freely available songs (tracks) with their audio features	Songs (tracks) released in 1902–2017
Spotify	Music Streaming Sessions Dataset (MSSD)	Session-based log data on music choices, tracks, and their audio features	July 2018–September 2018

Notes: iSpot.tv (<https://www.ispot.tv>); adland.tv (<https://adland.tv>); FMA (<https://freemusicarchive.org/>); MSSD (<https://doi.org/10.1145/3308558.3313641>).

television seasons from September 2015 through August 2018. The data set consists of over a hundred networks and covers all DMAs in the continental United States. Given the large number of ads over the three years, we limit our investigation to ads on five national broadcast networks: ABC, CBS, Fox, NBC, and The CW. Further, we consider only English-language programs and paid ads. Thus, ads by the networks promoting their own programming are not included.

Our unit of observation is a “creative insertion,” where a “creative,” or an “ad creative” or an “ad,” is a particular ad execution for a brand, and an “insertion” refers to the unique network, program, date, and time combination during which an ad was inserted. The Geico brand, for example, has several ad creatives. A 30-second ad for this brand is thus a different creative from a 15-second ad for the same brand. The same creative could also have multiple insertions over different networks, different programs, different dates, and different times. Our data also include the actual video of each creative. In addition, several variables coded by iSpot.tv are available. These variables include descriptors such as the emotion or emotions in the content, presence/absence of animals, and inclusion of popular music. Table 2 summarizes the metadata available for each ad.

The data selected on the basis of our aforementioned criteria include more than 1 million insertions of over 27,000 ad creatives by 3,200 brands across 15 broad product categories.⁷ Specifically, the data set records (1) the number of tracked TV sets tuned in for at least 3 seconds after the start of the ad (“start TVs”) and (2) the number of TV sets still tuned in when at least 25%, 50%, or 75% of the ad has played (“end TVs”). Only a monitored TV that is counted among the start TVs is eligible to be counted among the end TVs. For example, when an ad creative that is 30 seconds long is aired, all monitored TVs that played the ad during the first 3 seconds would be the number of start TVs, and the number

Table 2. Ad-Creative Metadata.

Variable	Description
Brand	Brand name associated with the ad creative
Duration	15 seconds (baseline), 30 seconds, 60 seconds, 90 seconds, >90 seconds
Promotion	Indicator for whether the ad creative includes a sales promotional message
Animal	Indicator for whether the ad creative displays animals
Song	Indicator for whether the ad creative has an accompanying popular song
Mood	<ul style="list-style-type: none"> • Active (baseline): Indicator for whether the ad creative is action-oriented • Emotional: Indicator for whether the ad creative is emotional • Informational: Indicator for whether the ad creative is informational • Funny: Indicator for whether the ad creative is humorous • Sexy: Indicator for whether the ad creative has a sexual theme

that continued to play it for at least 75% of the length (23 seconds) would be the end TVs. Thus, the difference between the number of start TVs and the number of end TVs reflects the number of TV sets that interrupted the ad’s showing before 25%, 50%, or 75% of its duration. Using the data set, we compute a measure of ad viewing called the ad-tuning rate as follows:

$$\text{Ad-tuning Rate} = \frac{\text{End TVs}}{\text{Start TVs}} \times 100. \quad (1)$$

Note that the “start TV” and “end TV” measures reset for every ad insertion. Suppose the first ad break of a program contains three ads. A television has to be tuned in to the ad for at least 3 seconds for the brand and ad creative to be identified. Thus, ads that are interrupted instantaneously are not included in the measurement. As an example, consider a 30-second ad. If 1 million televisions passed the 3-second threshold for the first ad in the ad break and 720,000 televisions are tracked at the

⁷ The 15 broad product categories coded by iSpot.tv are apparel, footwear, and accessories; business and legal; education; electronics and communication; food and beverage; health and beauty; home and real estate; insurance; life and entertainment; pharmaceutical and medical; politics, government, and organizations; restaurants; retail stores; travel; and vehicles.

23-second mark, which is 75% of the length of the ad, then the ad-tuning rate at the 75% cutoff for this ad is $720,000/1,000,000 = .72$, or 72%. When the second ad is aired, the tracking starts afresh. Suppose this ad is also 30 seconds long, but this time 1.2 million televisions are tuned in for the first 3 seconds of this ad and 950,000 televisions are still tuned in at the 23-second mark. The ad-tuning rate for this ad is thus $950,000/1,200,000$ or 79%. The second fact to note is that TVs that tune in to the ad *after* the first 3 seconds are not counted. Only televisions that are on for the first 3 seconds of the ad are tracked. Thus, the ad-tuning rate measure is specific to each ad and is unaffected by the number of TVs tuned in before or after the ad in question.

For the 75% cutoff, the mean ad-tuning rate across insertions of all the ad creatives in our data is 82.67%, and the standard deviation is 18.80%. At the 50% cutoff, the mean value is 94.33% with standard deviations of 10.48%, and at the 25% cutoff, the mean is 99.34% and the standard deviation is 3.53%. That is, the ad-skip rates in our data range between 5.67% and 17.33%, for cutoff values of 50% and 75%, respectively. These numbers may seem low, but they are not too far from those reported in previous studies. For instance, Danaher (1995) finds that television ratings drop by about 5% during commercial breaks in a data set from New Zealand. More recently, Tuchman, Nair, and Gardete (2018) report an ad-skip rate of about 5% in a data set from a Western European country. Using a TiVo data set from the United States, Deng and Mela (2018) report an ad-skip rate of about 15% for live viewing. Nonetheless, we provide some reasons for why the tuning rates in our study are higher than commonly held priors that suggest lower ad watching. First, those who immediately switch out within 3 seconds are excluded from the base because the ACR software needs this time to identify the ad creative. In particular, this exclusion partly explains why the view rate exceeds 99% with the 25% cutoff. For instance, the 25% duration of a 15-second ad is 3.75 seconds. Second, our measure omits a portion of time-shifted views using a DVR, for which ad skipping is more frequently observed (e.g., Deng and Mela 2018). Third, the measure does not account for the population that watches TV programs via streaming. Because we use the same measure for all the ads that we investigate, however, these reasons will not distort our findings.

We use three additional data sets for our investigation:

1. **Free Music Archive:** The FMA is a publicly available data set used widely in the music information retrieval (MIR) literature (Defferrard et al. 2017). We use these data to develop an approach to reverse engineer Spotify's proprietary audio feature measurement methodology.
2. **Super Bowl:** We collected the videos of all 3,077 ad creatives that aired during the Super Bowl between 1969 and 2020 from the website adland.tv. We use these data to explore the changes in the energy level of these commercials over a long period.

3. **Music Streaming Sessions Dataset:** As of 2019, the MSSD is the largest publicly available data set for researchers to track consumers' preferences in streaming music and includes details of the musical tracks played and the specific hour and minute of the day at which the tracks are played (Brost, Mehrotra, and Jehan 2019). The data set includes a log of a total of 3.7 million tracks that were played in 150 million listening sessions on Spotify, and we use a 20% random sample for computational tractability. We use these data to characterize the temporal patterns in the energy levels of streamed audio tracks.

Measuring Energy in Ad Content

The empirical question of interest is whether the association between the level of energy in television ads and the extent of ad tuning is positive or negative. To examine this association, we measure the energy level of each ad in our data set. In this section, we explain our operational measure of energy in ads using the Spotify audio energy measure. We then present data patterns from several descriptive analyses.

An Operational Measure of Energy in Ads

To measure the audio energy in TV commercials, we draw on work done by a company called Echo Nest, which was founded in 2005 as a research spin-off from the MIT Media Lab and acquired by Spotify in 2014. Echo Nest developed a proprietary approach to measure multiple characteristics of audio tracks (hereinafter Echo Nest or Spotify attributes), which some consider "the current gold standard in MIR" (Askin and Mauskopf 2017). The Echo Nest attributes include seven subjective measures, which are labeled acousticalness, danceability, energy, liveness, instrumentalness, speechiness, and valence, and one objective measure, namely, tempo.

Developers and researchers can use the Spotify application programming interface (API) to get the Echo Nest attributes of tracks that are available in Spotify's music database. However, we cannot use Spotify's API to estimate the audio energy in the TV ads because TV ads are not listed in the Spotify database.⁸ Moreover, our focus is on the *overall* auditory energy of ads, where the audio includes not only background music but also other sounds (e.g., speech and nonmusical sounds). Therefore, we develop an alternative approach that relies on an open-source algorithm called Librosa (McFee et al. 2015) and the FMA data set mentioned previously.

Librosa can be used to decompose audio into a large number (>500) of low-level spectral and rhythmic audio primitives or

⁸ In our data set, 85% of the TV ads do not use music from tracks that are available on Spotify, and the other 15% contain only snippets of the songs, which precludes the direct use of the Spotify API even if our focus is solely on background music. The use of recorded music in TV ads involves paying copyright fees, which is the reason that very few TV ads use such music.

Table 3. Example of TV Ads with Highest and Lowest Energy Levels.

	Highest Energy	Lowest Energy
1	Under Armour TV commercial, “We Will” featuring Michael Phelps, Misty Copeland	Nike TV commercial, “Until We All Win” featuring Serena Williams
2	Colgate TV commercial, “Every Drop Counts”	Center for Biological Diversity TV commercial, “Polar Bear”
3	OMEGA Seamaster 300 TV commercial, “Spectre: Revealing the 007 Watch”	TaylorMade TV commercial, “Expect the Unexpected”
4	Fruit of the Loom TV commercial, “Holidays: Feel Free to Celebrate”	Clorox Bleach TV commercial, “Bleach It Away: Distance”
5	Ford TV commercial, “We Are All Champions”	ASPCA TV commercial, “Baxter”
6	McDonald’s McCafé TV commercial, “Nothing Before Coffee: Downpour”	Ralph Lauren Fragrances Tender Romance TV commercial, “Love”
7	Apple Music TV commercial, “Apple Music Anthem” song by Noga Erez	Windex TV commercial, “The Story of Lucy: Just the Beginning”
8	2018 Honda Accord TV commercial, “Tower of Success”	PNC Bank TV commercial, “Know You’re Saving for Special Moments”
9	M&M’s Super Bowl 2018 TV commercial, “Human” featuring Danny DeVito, Todrick Hall	Xeljanz TV commercial, “Birthday Puppy”
10	Coca-Cola TV commercial, “Final Four: One Last Dance”	Blue Apron TV commercial, “Farm Fresh Ingredients”

Notes: The table lists examples of TV commercials with the highest and lowest values of energy in our data set. The URL links to the videos of these ad creatives can be found in Web Appendix A.3.

features. These features include chroma, mel-frequency cepstral coefficients, root-mean-square energy, spectral bandwidth, spectral contrast, tonnetz, zero crossing rate, and others, for which statistics such as minimum, maximum, median, mean, standard deviation, kurtosis, and skew are reported. The FMA data set includes 13,129 tracks for which both the Librosa audio features and Echo Nest audio attributes are available.

Using the data set, we reverse-engineer the Echo Nest attributes by building a model that predicts the Echo Nest attributes based on the Librosa features. We communicated with a developer of the Echo Nest attributes, asking if it would be appropriate to derive Echo Nest attributes from the Librosa features. The developer responded that (1) a prediction model using Librosa features could work in principle, and (2) it would be an approximation of the original model. We summarize the process next (see Web Appendix A for more details):

1. Using the FMA data set, we employed machine learning methods to derive the relationship between the Echo Nest energy attribute and the Librosa features. We explored different model forms (e.g., nonlinear) and machine learning methods (e.g., deep learning) to choose the model that achieved the best out-of-sample prediction.
2. We extracted the audio component of each ad creative in our iSpot.tv data using FFmpeg, which is an open-source project that can extract the audio from all types of video files, such as TV ads.
3. We used Librosa to decompose the audio files from Step 2 into the 518 low-level spectral and rhythmic features.
4. We used the trained machine learning models (from Step 1) to estimate the energy level of each ad creative using the Librosa features from Step 3.

By doing so, we implicitly assume that the relationship between the Librosa features and the Echo Nest attributes from the FMA data set holds for audio in general. We do this not only for energy but also for the other Echo Nest attributes. While the prediction precision varies across attributes, we are confident in reverse engineering the energy attribute. For instance, the out-of-sample R-square value for energy is about .8, and the root mean square error for this attribute is the lowest among all the attributes.

Table 3 presents sample TV ads with the highest and lowest predicted audio energy levels (see Web Appendix A.3 for URL links to the commercials). In our own evaluation, the two sets of ads are substantially different in terms of how energetic we perceived the commercials to be, with some exceptions (e.g., the Colgate TV commercial is not particularly energetic). Of course, our evaluation is subjective. To better understand what our measure of energy represents, we need a more systematic approach, which we discuss in the next section.

Data Patterns

We present results from three exploratory data analyses that describe the variation of the auditory energy level in ads across time, product categories, and program genres. These analyses document the following patterns using our data sets: (1) The energy level of TV ads is growing year after year with distinct intraday and intraweek temporal patterns. (2) Energy levels are heterogeneous across product categories. (3) Energy levels vary substantially across ads placed in different genres of TV programs. These findings are discussed in this section.

Energy levels over time. Figure 1, Panel A, depicts the average monthly energy levels of all ad creatives from the iSpot.tv

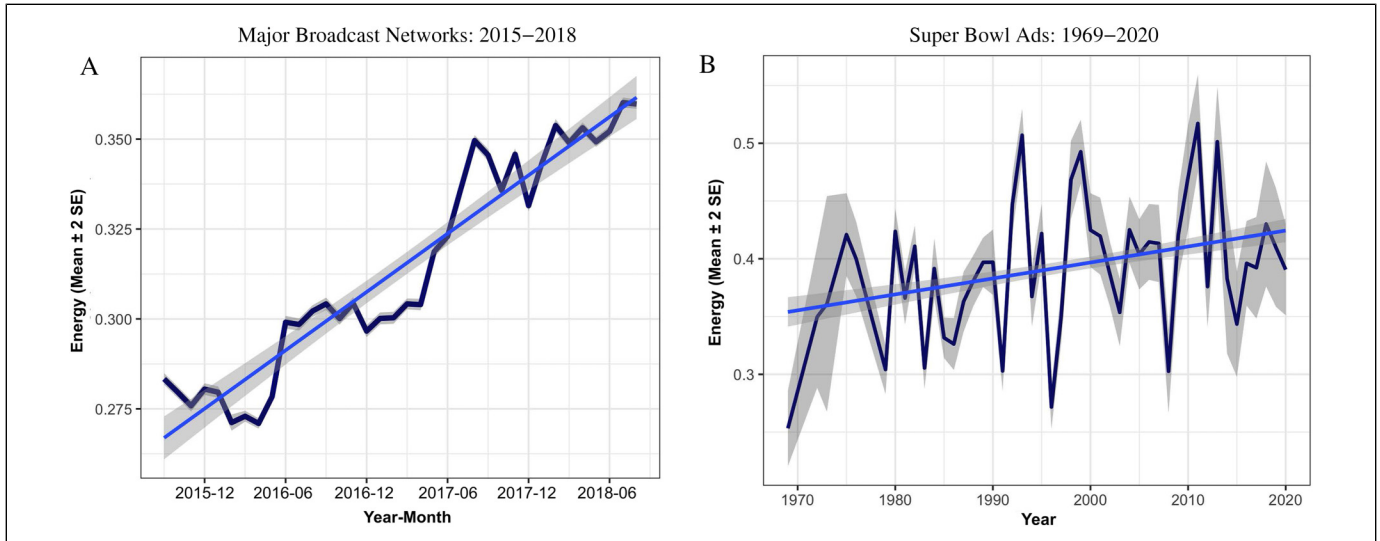


Figure 1. Trends in energy levels of TV commercials.

Notes: The two panels show that the average energy level of TV ads is growing year over year: Panel A illustrates the average auditory energy level of the ads shown on the five major broadcast television networks (ABC, CBS, FOX, NBC, The CW) in the iSpot.tv data set (September 2015–August 2018). Panel B demonstrates the average auditory energy level of the Super Bowl ads from Super Bowl III (1969) to Super Bowl LIV (2020).

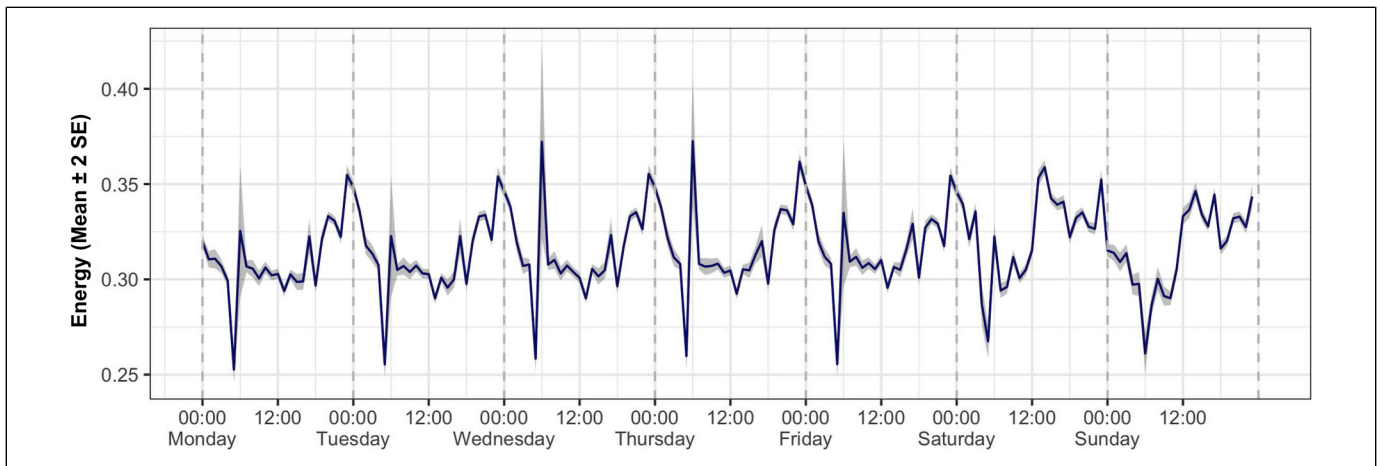


Figure 2. Temporal patterns in the energy levels of TV commercials.

Notes: The figure displays the temporal patterns in the auditory energy level of the ads shown on the five major broadcast television networks (ABC, CBS, FOX, NBC, The CW) in the iSpot.tv data set (September 2015–August 2018). The solid line represents the average energy level of all ad creatives for each hour of the day.

data set, showing that the energy levels are increasing over these three years. The linear fitted value of the energy level is .278 in September 2015 and .367 in August 2018, which indicates a 33% increase during that period. To further investigate the temporal pattern over a longer horizon, we extracted the audio component from Super Bowl ads spanning six decades, and we plot their energy levels in Figure 1, Panel B. Each year includes ads with high energy levels and ads with low energy levels, as seen in the dispersion, but there is a clear increasing trend in energy levels over the years. The linear fitted value of the energy level in 1969 is .354, and that in 2020 is .424, which indicates a 20% increase during the time

period. One caveat is that Super Bowl ads are unique because advertisers tend to employ creative executions that are more dramatic than usual, as seen in the higher energy levels for Super Bowl ads.

Next, Figure 2 illustrates the change of energy levels in the TV ads by hours within a week using the iSpot.tv data set. The temporal patterns are distinct: the energy level dips around 5 A.M. and peaks around 11 P.M. The level of some ads rallies around 6 A.M. on weekdays, but the phenomenon does not hold on weekends. In general, commercials on weekend afternoons (12 P.M.–6 P.M.) are more energetic than those aired on weekdays.

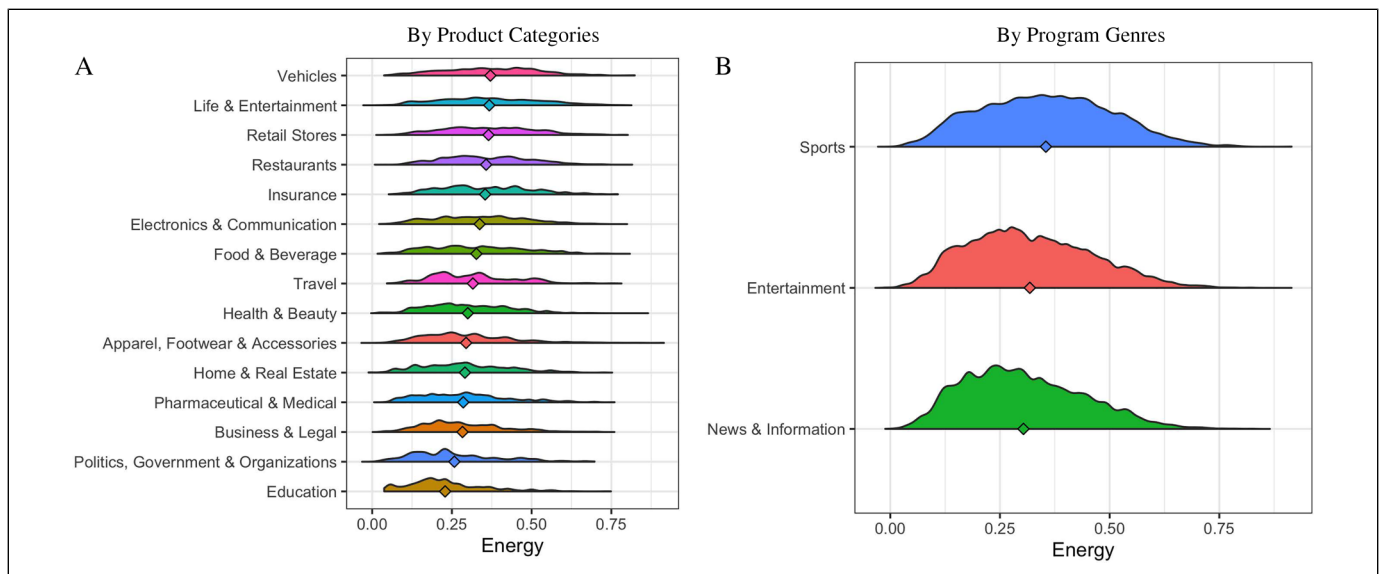


Figure 3. Energy levels of TV commercials by product categories and program genres.

Notes: Panel A plots the distributions of energy level for each product category. The product categories are ranked based on their mean energy levels. Panel B displays the distributions of energy level for ads placed in each program genre.

Energy levels across product categories. As shown in Figure 3, Panel A, energy levels in ads vary across the 15 product categories in the iSpot.tv data set. The product categories are ranked on the basis of their mean energy levels. The top three are vehicles (mean = .371), life and entertainment (mean = .367), and retail stores (mean = .365), while business and legal (mean = .283), politics, government, and organizations (mean = .258), and education (mean = .229) make up the bottom three. Typically, ads for trucks, sport-utility vehicles, entertainment, and retailers tend to be fast, loud, and noisy. In contrast, the bottom three categories, which typically use somber and informational creatives, have low energy. Specifically, ads for business and legal services as well as those for educational institutions and governments are low on energy. The mean differences between the top and the bottom three categories are statistically significant ($t = 156.56, p < .001$).

Energy levels across program genres. The three main program genres are sports, entertainment, and news and information. Two other genres in the data set, labeled “various programs” and “infomercial,” are dropped in this and subsequent analysis because their sample sizes are negligible. Figure 3, Panel B, displays the distributions of energy levels for ads placed in these three genres. Ads aired during sports programs have the highest energy on average (mean = .355), followed by ads aired during entertainment programs (mean = .318). In contrast, the genre with the lowest energy on average are ads placed in news and information programs (mean = .304). The mean differences are statistically significant for sports versus entertainment ($t = 70.79, p < .001$) and news and information ($t = 90.84, p < .001$). Further, entertainment is significantly different from news and information ($t = 45.38, p < .001$). We expect sports programs to be more competitive and intense, whereas

news broadcast programs are less upbeat, and entertainment programs will fall somewhere in between. The fact that the energy levels of TV ads in our data set vary in the same order across program genres indicates that advertisers may match their ads with the energy level of TV programs.

Understanding Energy in Ad Content: Algorithm-Generated Versus Human-Perceived Energy

What does Spotify ad audio energy represent? To answer the question, we borrow the arousal–valence framework developed by Russell (1980). The model provides a mapping of emotions onto a two-dimensional space, where the two axes are defined as arousal and valence that share the same aforementioned definitions from Di Muro and Murray (2012). For instance, excited, delighted, and happy emotions are mapped into the first quadrant as they are high on arousal and the valence is positive. Alarmed, afraid, and angry are in the second quadrant, which is high on arousal and has negative valence. On the other hand, sad, tired, and bored fall into the third quadrant as they are low on arousal and the valence is negative. Finally, emotions such as calm, relaxed, and contented are mapped in the fourth quadrant, being low on arousal with positive valence.

Our goal is to understand how our algorithm-based, objective measure of energy in ad content is related to the subjective measures of affect along the arousal and valence dimensions. Specifically, we compare human-perceived energy in ads evaluated by Amazon Mechanical Turk (MTurk) participants to the Spotify audio energy estimated for these ads on the two dimensions. We report a full description of our MTurk study in Web

Appendix B. Here, we briefly explain the measurement and the results.

The measurement scales used in the MTurk study were borrowed from Puccinelli, Wilcox, and Grewal (2015) and Belanche, Flavián, and Pérez-Rueda (2017). We used multiple items with seven-point bipolar scales to assess arousal and valence. Items 1–7 (“Not energetic” to “Energetic,” “Dull” to “Exciting,” “Not animated” to “Animated,” “Inactive” to “Active,” “Relaxed” to “Stimulated,” “Calm” to “Excited,” and “Unaroused” to “Aroused”) were used to load on the arousal factor, while items 8–12 (“Unhappy” to “Happy,” “Displeasure” to “Pleasure,” “Feel bad” to “Feel good,” “Sadness” to “Joy,” and “Negative” to “Positive”) were used to load on the valence factor. We reverse coded some of the items to prevent straightlining and randomized the order of the items to guard against acquiescence bias. The internal consistencies (Cronbach’s alpha) were all above .95 for both arousal and valence.

We collected data from 2,342 MTurk participants on a randomly selected sample of 138 ad creatives.⁹ Participants were randomly assigned to one of three conditions: (1) they can only hear the audio of ads without video (audio-only), (2) they can only watch the video of ads without audio (video-only), or (3) they can hear and see both audio and video of ads (audio + video). By doing so, we aim to understand how similarly or differently the MTurk participants perceive the levels of arousal and valence from different stimuli in ad content (i.e., auditory, visual, or both).

Overall, we find that the Spotify audio energy is related to the level of arousal in ad content. The correlations between the Spotify audio energy and human-perceived arousal in ad content from the audio-only, video-only, and audio + video conditions are .44, .38 and .46, respectively, in the main study (Figure W2 in the Web Appendix). We cannot say whether these correlations are high, but importantly, Spotify audio energy is consistently correlated with multiple measures of arousal (e.g., energetic vs. nonenergetic; exciting vs. dull; active vs. inactive; stimulated vs. relaxed; aroused vs. unaroused; see Figures W4 to W6 in the Web Appendix). In contrast, the correlation between Spotify audio energy and valence from the audio-only, video-only, and audio + video conditions are lower at .14, .21, and .13, respectively, in the main study (Figure W3 in the Web Appendix). This finding is reassuring as the correlation between energy and valence derived using our machine learning algorithm from the FMA data set is .22 (Table W6 in the Web Appendix).

In addition, the correlations between the levels of perceived arousal from the three conditions are .57 between audio-only and video-only, .74 between audio-only and audio + video, and .86 between video-only and audio + video. We believe

this high level of correlation is due to close alignment between the arousal stimuli in the audio and video in ads. The high correlation in the levels of arousal from the audio-only and audio + video conditions suggests that the Spotify energy measure can capture the overall arousal level and not just that from the audio in ad content.

We also obtain additional insights on the human-perceived arousal in ads. For instance, we find that Spotify audio energy is significantly more correlated with the arousal in background sound than with the arousal in human speech in the audio-only condition (.53 vs. .33; $z = 2.616$; $p = .009$) and the audio + video condition (.48 vs. .35; $z = 1.691$; $p = .091$; see Figure W7 in the Web Appendix). This finding is not particularly surprising, because Echo Nest originally designed the algorithm to summarize the attributes of music rather than speech.

Additional reinforcing evidence of the positive correlation between energy and arousal comes from the keywords provided by survey participants when they were asked to “list some words that come to mind when you think about energetic ads” (see Web Appendix B.1.5). The top five keywords mentioned by participants in the audio + video condition were “fast,” “music,” “movement,” “upbeat,” and “exciting.” In the audio-only condition, participants listed “music,” “upbeat,” “fast,” “loud,” and “exciting,” whereas in the video-only condition, they listed “movement,” “fun,” “active,” “fast,” and “exciting” (Figure W9 in the Web Appendix). The emotions represented by the keywords appear to fall in the high arousal–high valence quadrant. Note that the keywords from the audio + video condition appear to be the union of the keywords in the audio-only and video-only conditions. The results clearly suggest that both auditory and visual elements contribute to the overall energy level in ads, as measured by arousal. In Web Appendices B.2 and B.3, we further investigate auditory and visual correlates of energy in ad content.

Associations Between Ad Energy and Ad-Tuning Rate

In this section, we explore the association between the ad-tuning rate and the energy level of ad content. The dependent variable of interest is the ad-tuning rate, ATR_{it} , measured as the percentage of TVs tuned in to an ad for at least 75% of its duration (we also discuss the results from the 50% and 25% cutoff values). The index i represents a unique creative execution of an ad, and the index t represents the insertion, which is a unique combination of network, program, date, and time. The independent variable of interest is $Energy_i$, which is the energy level of an ad creative computed using the first 75% of its duration. This variable accounts for the fact that for a tuning rate measured for a 75% cutoff, viewers would have been exposed to the ad’s energy during the first 75% of its duration.¹⁰

⁹ We randomly selected 30 ad creatives for a pilot study and 110 ad creatives for the main study. Two of the 30 ads in the pilot study were excluded because of a copyright issue regarding the music that accompanied the ads. The cost was about \$2,500 to rate the energy level of 138 ads through MTurk, so evaluating all 27,000 ads would have cost at least half a million USD.

¹⁰ The correlation between the energy levels measured using the first 75% and 100% of ads’ duration is about .962. The results are qualitatively unchanged when 100% of ad duration is used.

To recover the association between ATR_{it} and the energy level of ad creative i , $Energy_i$, we employ two empirical strategies: a between-estimator approach and a within-estimator approach.

A Between-Estimator Approach

Here we investigate the association between ATR_{it} and $Energy_i$ using a two-stage model. In the first stage, we use a fixed-effects model to quantify the ad-creative fixed effects (denoted by δ_i), which we interpret as an insertion-invariant measure of the “quality” of ad creative i in terms of ad tuning. Specifically, the ad-creative fixed effects capture all the components that constitute an ad creative *and* explain audiences’ ad-tuning behavior. In the second stage, we project various aspects of ad creative i , including $Energy_i$, to the estimated δ_i from the first stage. Next, we explain the specification and the results of each stage.

First stage: Measuring ad-creative fixed effects. We first fit the following model:

$$ATR_{it} = \delta_i + X_t'\gamma + e_{it}, \quad (2)$$

where δ_i represents the ad-creative fixed effects, X_t is a set of variables that summarizes the insertion of i , and e_{it} is an error term. For X_t , we include network fixed effects, program fixed effects, year-week fixed effects, day-of-week fixed effects, day part fixed effects,¹¹ and position fixed effects. Network fixed effects are for the five broadcast networks, while program fixed effects are specific to each program, such as *The Big Bang Theory* (CBS) or *This Is Us*—(NBC). Multiple networks carry some events, such as presidential debates and political conventions, and they also occasionally rebroadcast other networks’ programs, which allows us to identify both network and program fixed effects. “Position” refers to the pod and pod position in which the ad is shown. A program consists of several ad breaks called pods, and each pod has several ads represented by their position in the pod—for example, the first ad in the first pod, the third ad in the second pod, and so on. Previous studies have demonstrated that viewing behavior can systematically differ across the positions of ads within a program (e.g., Danaher 1995; Van Meurs 1998). We include a fixed effect for each pod and for each position within each pod, collapsing pods after the tenth position at ten.

To assess the relative ability of various fixed effects to explain ATR_{it} , we compute the variation in ATR_{it} explained by our fixed effects when used either alone or in a combination. To ensure reliable estimation of the ad-creative fixed effects, for this part of the study, we drop about 8,200 ad creatives with fewer than

five insertions in total during the three-year period. These ads account only for 1.6% of insertions in our data.¹² We find that ad-creative fixed effects alone explain 55% of the variation, which is far greater than that explained by network fixed effects (.17%); network and program fixed effects (2.20%); year-week, day of week, and day part fixed effects (2.59%); and network, program, year-week, day of week, and day part fixed effects (4.45%). It is logical that the strongest explanation of ATR_{it} is provided by the ad-creative fixed effects. When all the fixed effects are used, we explain about 57% of the variation in ATR_{it} . We report this variation for all possible combinations of fixed effects in the Web Appendix (Tables W11 and W12).

We use weighted least squares with the square root of the number of viewers as weights to estimate the full regression in Equation 2 (i.e., the variance of the observed ad-tuning rate is assumed to be inversely proportional to the number of viewers). The use of these weights accounts for the relative precision of ad-tuning rates that are measured from different numbers of viewers. The specification of the regression thus provides estimates of the ad-creative fixed effects while controlling for the network and program on which the ad is shown, and time and ad position. Higher δ_i indicates that ad creative i is intrinsically more likely to be tuned in to. Figure 4, Panel A, presents the distribution of the estimated δ_i values and shows that there is substantial variation in the estimated ad-creative fixed effects, which we exploit in the next stage. The estimates of other fixed effects are reported in Web Appendix C.

Second stage: Decomposition of ad-creative fixed effects. As explained, the estimated ad-creative fixed effects capture all the components that constitute an ad creative (e.g., brand, the message, and ad duration) *and* explain audiences’ ad-tuning behavior. In Figure 4, Panel B, we plot the distribution of the estimated ad-creative fixed effects and the corresponding estimated energy levels of the ads and find a positive relationship. We next formally investigate whether there is still a meaningful association between the energy level of ads and ad-tuning rates even after controlling for other factors of the ad. Specifically, we utilize the variables in Table 2 that measure the characteristics of the content of ad creatives and run the following regression:

$$\delta_i = \alpha + \beta \times Energy_i + Z_i'\eta + \varepsilon_i, \quad (3)$$

where Z_i includes all the variables in Table 2.

The dependent variable in Equation 3 is measured with error, very likely heteroskedastic, because it is an estimate from the first-stage regression. Following Lewis and Linzer (2005), we use a feasible generalized least squares (FGLS) method to account for the errors in the dependent variable and the model error in Equation 3.¹³ We provide more details on this

¹¹ A 24-hour day is divided into nine day parts on television: early morning (Monday–Friday 6–10 A.M.), daytime (Monday–Friday 10 A.M.–4 P.M.), early fringe (Monday–Friday 4–8 P.M.), prime time (Monday–Saturday 8–11 P.M., Sunday 7–11 P.M.), late fringe P.M. (Monday–Sunday 11 P.M.–12 A.M.), late fringe A.M. (Monday–Sunday 12–2 A.M.), overnight (Monday–Sunday 2–6 A.M.), weekend day (Saturday–Sunday 6 A.M.–1 P.M.), and weekend afternoon (Saturday 1–8 P.M., Sunday 1–7 P.M.).

¹² The results are robust when we drop ad creatives with fewer than 10, 20, or 30 insertions in total, although the reduced sample size affects the precision of estimates.

¹³ An alternative approach is to use a hierarchical model in which we combine Equations 2 and 3 into a single unified model with a random effect, rather than having fixed effects. This approach could result in a more accurate estimation of the standard error for energy.

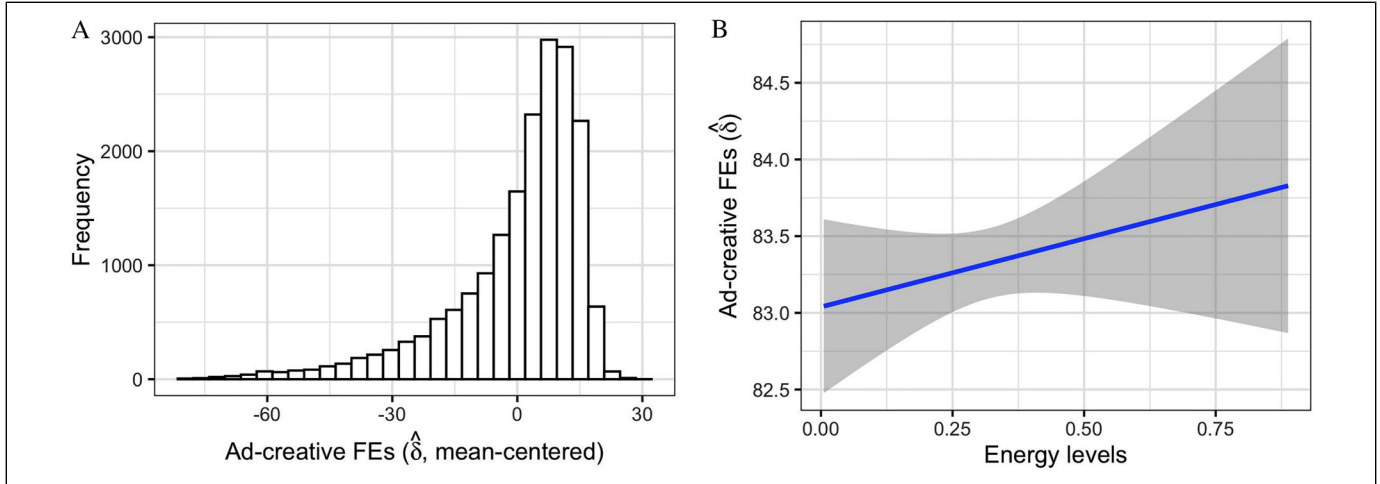


Figure 4. Estimated ad-creative fixed effects.

Notes: Panel A reports the distribution of ad-creative fixed effects (mean-centered) from the first-stage regression. Higher $\hat{\delta}_i$ indicates that ad creative i is intrinsically less likely to get avoided. Panel B reports the association between estimated ad-creative fixed effects and energy-level. The solid blue line is the regression line, and the gray band displays the confidence interval.

procedure and the estimation results from other approaches in Web Appendix D.

Table 4 reports the estimation results of Equation 3. The “brand” in Table 2, which is the brand that was being advertised in the ad creative, is modeled using brand fixed effects. It is clear from the improvement in fit in Table 4 that the model would be misspecified without controlling for the brand. We find that the energy level is positively associated with the ad-creative fixed effects, and the association is statistically significant when we include brand fixed effects. Thus, ad creatives with higher energy levels will have larger fixed effects in Equation 2, which, in turn, means that they will have higher tuning rates as well.¹⁴

In terms of the magnitude of the estimate, we find that an increase of one standard deviation in an ad creative’s energy level (.137) is associated with a $.137 \times 3.343 = .458$ percentage point increase in the ad-tuning rate. The number translates into about 5,079 more TVs tuned in for 75% or more of the ad’s length per insertion (based on an average of 1,108,957 TVs tuned in per ad insertion in our data). Note that over 200 brands have within-brand energy level standard deviation greater than .137 (e.g., The North Face with .282, Apple Music with .246, and Under Armour with .246), which suggests that such changes in energy level are within the realm of possibility for advertisers. Overall, these results provide support for

the positive association, namely, that more energetic ad content results in longer ad-tuning rates.

We also detect interesting patterns in the other coefficients. For instance, longer-duration ads have lower ad-creative fixed effects and thus would have lower ad-tuning rates. When compared with the baseline duration (i.e., 15-second ads), 90-second, 60-second, and 30-second ads, in that order, are less likely to be tuned in for more than 75% of their length, which makes intuitive sense. This pattern is also consistent with previous findings that longer-duration ads are more likely to be avoided (Woltman Elpers, Wedel, and Pieters 2003, p. 445). Our model-free raw data also show that the average ad-tuning rate systematically declines with increasing length of the ad. The results also indicate that (1) sales promotional ad creatives or ads with accompanying popular songs are less likely to be watched, (2) ad creatives that display an animal or animals are more likely to be watched, and (3) informational ad creatives are less likely to be watched than active ad creatives (active is the baseline for the mood variable). A detailed explanation of these findings is beyond the scope of this research.

A Within-Estimator Approach

The model results discussed previously relied on the between-ad variation in energy levels. That is, we compared ad A and ad B with different energy levels while controlling for the specificities of an insertion (via all the insertion fixed effects) and some other aspects of ad content. One concern regarding the between-estimator approach is potential omitted variable bias as one cannot test the sufficiency of a given set of control variables. Here, we propose an alternative estimator that relies on “artificial” within-ad variation in energy, which we generate by exploiting the temporal patterns in audience preference for high-energy audio.

¹⁴ To account for a potential nonlinear relationship between energy and ad-creative fixed effects (Figure 4, Panel B), we also use alternative functional forms for the energy variable: a log transformation and a quadratic term of energy. We still find a positive and statistically significant relationship between ad-creative fixed effects and $\log(\text{Energy}_i)$. The quadratic term, however, does not work because of the high correlation of .97 between the linear and the quadratic energy terms.

Table 4. Between-Estimator Approach: Second-Stage Estimation Results.

	DV: Estimated Ad-Creative Fixed Effects ($\hat{\delta}_i$)			
	(1)	(2)	(3)	(4)
Energy	.928 (.811)	1.094 (.802)	1.180 (.811)	3.343*** (.922)
Duration: 30 seconds		-2.790*** (.228)	-2.877*** (.229)	-1.806*** (.278)
Duration: 60 seconds		-6.308*** (.546)	-5.939*** (.560)	-3.796*** (.789)
Duration: 90 seconds		-3.872** (1.657)	-2.378 (1.674)	-10.680*** (2.710)
Duration: >90 seconds		-21.842*** (1.774)	-20.998*** (1.789)	-21.305*** (2.317)
Promotion		-5.033*** (.258)	-4.954*** (.259)	-2.401*** (.369)
Animal		1.841*** (.389)	1.678*** (.390)	.940** (.479)
Song		-1.018*** (.298)	-1.151*** (.300)	-.647* (.357)
Mood: Emotional			1.427** (.673)	.465 (.791)
Mood: Informational			-2.105*** (.518)	-1.056* (.615)
Mood: Funny			.589** (.278)	.520 (.342)
Mood: Sexy			-.271 (1.279)	-1.848 (1.850)
Constant	83.269*** (.284)	86.181*** (.311)	86.156*** (.322)	
Brand fixed effects	No	No	No	Yes
N	18,933	18,923	18,861	18,861
R ²	<.001	.042	.044	.257
Adj. R ²	<.001	.041	.043	.152

* $p < .10$.** $p < .05$.*** $p < .01$.

Notes: The table reports the estimation results of Equation 3 using FGLS. Energy is measured using the first 75% of ads' duration. Standard errors are reported in parentheses.

The ideal variation should arise from randomizing the energy level of an ad creative while holding everything else fixed (i.e., insertion and other ad content). That is, we want to have *within*-ad variation in the energy level, where the source of variation is not correlated with other aspects of the ad. Using such variation, one can isolate the effect of energy levels in ads on outcome variables of interest from those held fixed. The challenge is that there is no such variation in the energy levels for a given ad creative due to the observational nature of our data.

To partially address the problem, we rely on temporal variation in people's preferences for energetic audio. For instance, Park et al. (2019) report that people generally prefer more intense music during the day than at night. We also find the same pattern by examining the energy levels of 31 million listening sessions on Spotify (the MSSD), which represent 2.3 million tracks (or songs) played. A plot of the energy levels by time of day is shown in Figure 5.

This data set suggests a way of operationalizing the within-ad variation in energy levels. For a given ad creative on TV, the ad's energy level could be higher or lower than the preferred musical energy at the time when it is aired. By exploiting this natural occurrence, we can compare the ad-tuning rate of the same ad in the two cases: when the energy level of the ad is either higher or lower than the level generally preferred by people when the ad is aired. We consider the baseline energy level from the Spotify listening sessions in the MSSD (Figure 5) for each one-hour time slot over a 24-hour day of the seven days in the week as the preferred energy level. We compute the distance between the energy level of the particular ad creative shown in a specific one-hour time slot and the preferred baseline energy level for that time slot as computed from the MSSD data set. We relate the ad-tuning rate to the distance between an ad's energy level and the baseline level.

Specifically, we introduce a variable, dist_{it} , which measures the distance between ad creative i 's energy level measured at

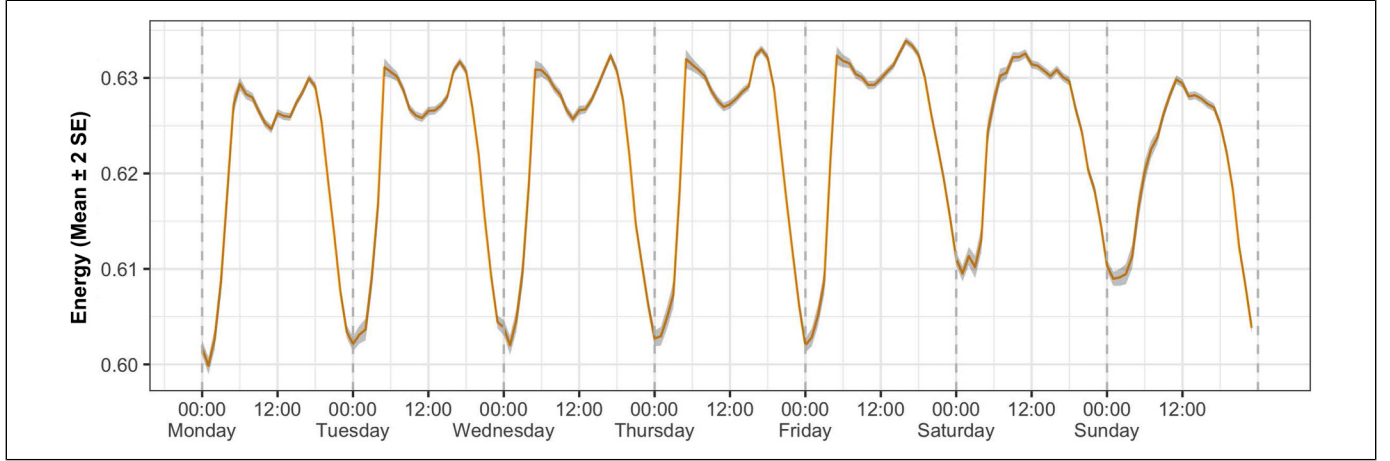


Figure 5. Temporal patterns in the energy levels of the MSSD.

Notes: The figure shows the temporal patterns in the auditory energy level of Spotify's MSSD. This figure was generated from analyzing the energy levels of 31 million listening sessions of 2.3 million songs between July and September 2018. During the weekdays, the energy level is relatively higher in daytime (6 A.M.–6 P.M.) than at night; it is highest at 6 A.M. and 6 P.M. and lowest at midnight. During the weekends, the energy level is highest around noon.

the first 75% of an ad's length (Energy_i) and the baseline energy level $\theta_{\tau(t)}$ (the subscript $\tau(t)$ indicates the day-hour pair of t ; e.g., Monday 1–2 P.M.). We cannot simply subtract the energy level of the TV ad creative from the energy level of the MSSD listening sessions because the level and variability of energy are different between the two. Therefore, we use deciles for both the energy level of ad creative i and the baseline energy level and compute the distance measure as $\text{dist}_{it} = \psi_i - \theta_{\tau(t)}$. Here, $\psi_i \in \{1, 2, \dots, 10\}$ is the decile of ad creative i 's energy level, and $\theta_{\tau(t)} \in \{1, 2, \dots, 10\}$ is the decile group of the preferred energy level at $\tau(t)$, computed using the MSSD data set. Thus, dist_{it} is one of the 19 integer values from -9 through 9 ; the more positive the value, the higher the relative energy level of ad creative i compared with the preferred baseline energy level.

Using our distance measure, we estimate the following model:

$$\text{ATR}_{it} = \delta_i + X_t' \gamma + \sum_{d \in \{-9, -8, \dots, 8, 9\}} \beta^d I[\text{dist}_{it} = d] + \varepsilon_{it}, \quad (4)$$

where δ_i are the ad-creative fixed effects and X_t includes the same set of fixed effects as in Equation 2, namely network, program, year-week, day-of-week, day part, and position fixed effects. The terms β^d are the parameters of interest.

We graphically report the parameter estimates in Figure 6. We find that a relatively higher energy level than the baseline is associated with higher ad-tuning rates or lower ad avoidance. The association is almost monotonic across different values of distance. Reassuringly, these results provide support for the positive association, i.e., that higher energy in ad content is associated with longer ad-tuning rates.

An advantage of using the Spotify music energy baseline is that it represents individual preferences for audio energy over the day. There are, however, several caveats with this approach.

First, the baseline is energy from Spotify music for a specific three-month period in 2018, whereas the ads in our data span a three-year period from 2015 to 2018. So, the temporal sample is not the same. Second, energy estimates for the MSSD listening tracks are directly provided by Spotify, whereas the energy estimates for TV ads were obtained with the machine learning procedure we described previously. Third, the baseline is music energy, but it is compared with ad energy, which is more than music. We cannot objectively confirm that this is a valid comparison, except to note that music is a major component of advertising. The fact remains, therefore, that Spotify listeners' preference for energetic music by itself may not be the same as the preferences of TV audiences and that the segments of TV audiences vary across different times.

To partially address these concerns, we use the aggregate energy level of ad creatives for a given time as the baseline.¹⁵ Specifically, we compute the baseline as the average energy level of ad creatives on a particular network in a particular month, day of week, and hour of the day. The underlying assumption here is that TV audiences have a prior expectation of the energy level of ads during a given time, which we summarize by the average energy level of ad creatives. In addition, advertisers select programs whose viewing demographics best match the product being advertised. Thus, one could expect average ad energy levels to reflect the desired energy levels. We find that an energy level higher than the baseline is positively associated with the ad-tuning rate, and the association is statistically significant, which supports our previous findings.¹⁶ For more details, see Web Appendix E.1.

¹⁵ We thank the review team for suggesting this idea.

¹⁶ Relatedly, we also use the energy level of the previous ad within the same commercial break as a baseline. We find that the estimate from this approach is statistically insignificant.

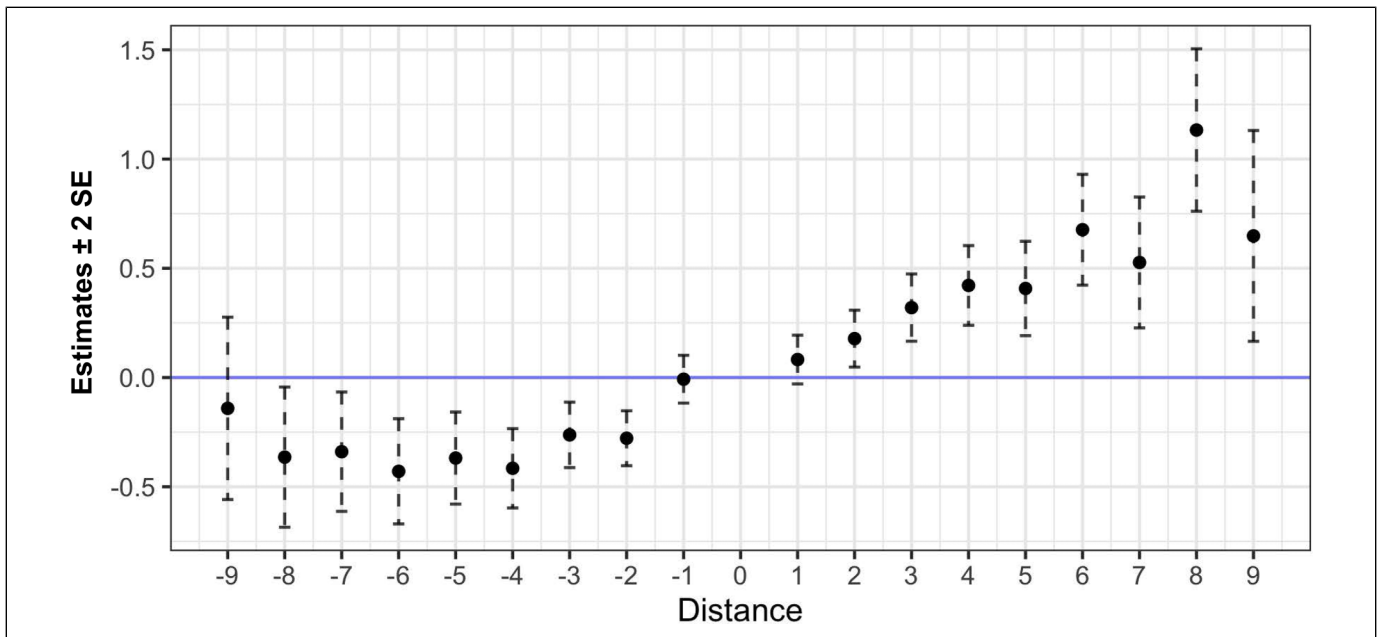


Figure 6. Estimation results of the within-estimator approach.
Notes: The figure reports the parameter estimates of β^d in Equation 4.

Interpretation

Overall, we find that the association between ad-tuning rate and energy levels in ads is positive and statistically significant using two empirical approaches. Nonetheless, we cannot claim that the association is causal, and we leave our findings as descriptive. This is because we cannot discount the possibility of an omitted variable bias as there could be missing variables that are related to the ad-creative effects (an example could be the message of ads). To the extent that these variables affect the ad-tuning rate and are also correlated with ad energy, our estimate is either under or overestimated. Further, given the observational nature of our data, the placement of ads with different energy levels is not randomized across programs and day parts, which together can create different contextual effects.

Relatedly, our reverse engineering of the Spotify algorithm using the FMA data set to measure energy in ad content could have generated a measurement error. The correlation between the predicted and the actual Spotify audio energy value available in the FMA data set is .89 in the holdout sample and .95 in the entire FMA data set. We find that the prediction is neither conservative nor extreme because most data points are aligned on the diagonal line without systematic bias (see Figure W1 in the Web Appendix). This measurement error could have biased the estimate toward zero, but we still find a positive and statistically significant association between energy in ad content and the average tuning rates. We expect that with a better prediction, for instance by having direct access to the Spotify algorithm, we would have been able to report a stronger association.

Lastly, because our energy measure is based on the auditory component of ads, it is worth noting that the association we

report may be intertwined with energy levels in the visual content of ads. As discussed previously, however, the energy levels of visual and auditory components of ads are highly correlated at least from the standpoint of human perception. Because energetic audio and energetic video tend to go hand in hand, one could measure the energy level of ads using either audio or video or both. Thus, our estimates can be interpreted as the *net* effect of energy levels in both audio and video of ads on ad-tuning rates.

Heterogeneity

In the model results presented previously, we estimated the average association between energy and the quality of ad creatives in terms of ad-tuning rates across all product categories and insertions. Here, we analyze whether and how the association between energy in ad content and ad-tuning rates varies across product categories and airings. Ideally, we would like to estimate β in Equations 3 and 4 for any given ad creative i and insertion t . In practice, however, this approach is infeasible because each airing of an ad creative is unique and, thus, some degree of aggregation is necessary. Toward this end, we estimate the association for combinations of the three program genres and 15 product categories in the data. The first column of Table 5 reports the estimation results of Equation 3 in which we interact the energy variable with the product category dummies. The next three columns give results for each of the three program genres, thus allowing us to quantify the ad-creative fixed effects specifically for each program genre. We estimate the first stage in Equation 2 separately for each program genre to obtain genre-specific ad-creative fixed effects.

Table 5. Heterogeneous Associations from the Between-Estimator Approach.

	DV: Estimated Ad-Creative Fixed Effects ($\hat{\delta}_{lig}$)			
	(1) All Programs	(2) Entertainment	(3) News and Information	(4) Sports
Energy				
× Apparel, footwear, and accessories	−5.821 (5.275)	−8.921 (7.257)	−12.384 (10.051)	−6.754 (8.591)
× Business and legal	8.719** (4.153)	1.070 (5.350)	7.114 (7.167)	9.659 (6.905)
× Education	3.073 (12.366)	−14.060 (17.152)	.229 (17.303)	8.693 (37.682)
× Electronics and communication	−.351 (2.280)	−1.004 (2.593)	7.775 (5.065)	−4.210 (3.800)
× Food and beverage	2.908 (2.200)	5.549** (2.476)	2.995 (3.353)	−4.740 (4.928)
× Health and beauty	.419 (2.741)	1.106 (2.864)	−1.842 (3.611)	5.824 (10.417)
× Home and real estate	1.242 (3.677)	1.480 (4.007)	−4.657 (4.936)	−3.478 (11.534)
× Insurance	8.608* (4.467)	8.050 (5.062)	9.791 (6.003)	1.587 (6.626)
× Life and entertainment	9.668** (3.963)	9.580* (5.160)	10.966 (11.181)	−4.065 (7.953)
× Pharmaceutical and medical	5.750 (5.227)	9.558* (5.393)	10.976** (5.616)	−16.058 (23.303)
× Politics, government, and organizations	11.072 (9.275)	10.843 (12.464)	11.483 (11.993)	15.715 (22.582)
× Restaurants	−.274 (3.625)	−1.489 (4.031)	2.516 (7.321)	−.014 (6.241)
× Retail stores	4.206 (2.718)	2.508 (2.954)	2.976 (3.890)	−18.993* (10.702)
× Travel	25.371*** (8.165)	34.235*** (9.879)	21.059 (14.016)	−18.145 (17.568)
× Vehicles	7.050** (3.219)	11.210*** (4.006)	−10.152 (7.266)	.346 (5.212)
Other controls ^a	Yes	Yes	Yes	Yes
Brand fixed effects	Yes	Yes	Yes	Yes
N	18,861	14,637	7,537	4,934
R ²	.259	.268	.295	.281
Adj. R ²	.152	.163	.148	.104

* $p < .10$.** $p < .05$.*** $p < .01$.

^a“Other controls” are the variables of duration, promotion, animal, song, and mood (see Table 2). The estimates for these variables under various model specifications are reported in the Web Appendix (Tables W22 to W25).

Notes: The table reports the estimation results of Equation 3 using FGLS by product categories and program genres. The sum of the sample sizes in each column exceeds the number of ad creatives because the same ad creative can be shown in entertainment, news and information, and sports genres. For news and information, network and day part fixed effects are dropped from the first stage because networks rarely rebroadcast news. Standard errors are reported in parentheses.

We find that the association of energy levels and ad-tuning rates is context dependent; that is, the magnitude and the statistical precision differ substantially across product categories and program genres. For instance, the association is statistically significant at the .05 level for only four product categories (business and legal, life and entertainment, travel, and vehicles) for all programs (Column 1). The number of categories with a significant relationship is even smaller when we consider a particular genre of programs: three for entertainment programs, one for news and information programs, and zero for sports programs.

One concern is that lack of statistical significance simply arises from the reduction in the sample size as we narrow our focus on a particular slice of the data. Consequently, we also check the results from a similar analysis using our within-estimator approach. In Table 6, we find that the association is statistically significant for more pairs of product categories and program genres, and some associations are even negative, although some estimates are different from Table 5, producing mixed results. Taken together, these results suggest that the boundary conditions for using high-energy features for TV commercials

Table 6. Heterogeneous Associations from the Within-Estimator Approach.

	DV: Ad Tuning-Rate (ATR_{it})			
	(1) All Programs	(2) Entertainment	(3) News and Information	(4) Sports
Energy				
× Apparel, footwear, and accessories	-.060 (.042)	-.010 (.101)	.230** (.110)	-.253* (.131)
× Business and legal	-.012 (.030)	.133** (.065)	.092 (.057)	.097 (.103)
× Education	-.003 (.078)	-.072 (.117)	-.086 (.192)	.129 (.373)
× Electronics and communication	.112*** (.023)	.125*** (.040)	.092 (.066)	.066 (.075)
× Food and beverage	.061*** (.020)	.242*** (.029)	.092*** (.035)	-.129 (.083)
× Health and beauty	.101*** (.021)	.281*** (.029)	.105*** (.035)	-.296** (.138)
× Home and real estate	-.001 (.025)	.170*** (.032)	-.112** (.048)	.306* (.165)
× Insurance	.119*** (.025)	.072* (.040)	.064 (.058)	-.157* (.086)
× Life and entertainment	.060 (.041)	.118 (.075)	.167 (.107)	.044 (.118)
× Pharmaceutical and medical	.045** (.022)	.234*** (.031)	-.014 (.035)	-.137 (.174)
× Politics, government, and organizations	.158** (.062)	.234* (.135)	.244** (.123)	.018 (.205)
× Restaurants	.060* (.032)	.025 (.053)	-.128 (.098)	.177* (.102)
× Retail stores	.113*** (.026)	.139*** (.038)	.145*** (.054)	-.006 (.138)
× Travel	.166*** (.045)	.142* (.082)	.350*** (.092)	.521*** (.174)
× Vehicles	.085*** (.028)	.145** (.057)	.255*** (.082)	.248*** (.087)
Ad-creative fixed effects	Yes	Yes	Yes	Yes
Insertion fixed effects ^a	Yes	Yes	Yes	Yes
N	1,057,798	728,246	237,267	83,484
R ²	.638	.653	.641	.711
Adj. R ²	.631	.645	.621	.675

* $p < .10$.** $p < .05$.*** $p < .01$.

^aInsertion fixed effects include network fixed effects, program fixed effects, year-week fixed effects, day-of-week fixed effects, day part fixed effects, and position fixed effects. For news and information (column 3), network and day part fixed effects are dropped because networks rarely rebroadcast news programs.

Notes: The table reports the estimation results of Equation 4 by product categories and program genres. For the ease of interpretation, the variable $dist_{it}$ is entered in the equation linearly. Standard errors are reported in parentheses.

will include (1) what products are advertised and (2) the types of programs in which the ads are placed.

Additional Analyses

We briefly discuss additional robustness checks for our findings. More details on each analysis are reported in Web Appendices E.2 and E.3.

Alternative definitions of the ad-tuning rate. Until now, we have used the 75% cutoff when defining the ad-tuning rate in the first

stage of our between-estimator analysis. We have two additional cutoffs at 50% and 25% available in the data. We cannot vary the cutoff as the data are already aggregated. Using the 50% and 25% cutoffs, we find that the estimate of the energy effect is not significant. One explanation for the lack of significance of energy when the ad-tuning rate is measured for earlier cutoffs could be that arousal takes some time to build, so the shorter the tuning time, the lower the effect. We find that the results are consistent when we transform the dependent variable (either arcsine or logit) in the first stage to alleviate the concern regarding the normality assumption of the ad-tuning rate, which ranges between 0 and 100.

In addition, we vary the cutoff values depending on the length of the ads. For 15-second or shorter ads, we used a 75% cutoff. For ads with length of more than 15 seconds, up to 30 seconds, we used 50%. For ads longer than 30 seconds, we used 25%. This approach allows us to keep the duration of initial exposure to ads roughly comparable before viewers take any actions of avoidance. We make sure that the energy levels of the ads are also measured by taking into account the varying cutoff values (e.g., the first 75% to measure the energy level of a 15-second ad, or the first 50% to measure the energy level of a 30-second ad). We find that the results are qualitatively unchanged.

Accounting for ad-wearout effects. The intensity of previous exposure to the same ad is likely to be an important determinant of individuals' ad-avoidance behavior (often referred to as "ad-wearout effects"). As we do not have individual-level viewership data, which would have made accounting for such an effect relatively simple, we construct a variable that represents cumulative viewership prior to a specific airing of an ad and use it as an additional control in the first stage of our between-estimator analysis. Because we do not know when an ad was first aired, we split our three-year data into two periods (the first 12 months and the remaining 24 months), and use the first period to address the initial condition problem and the second period to reestimate the model. The estimate for the cumulative viewership variable is negative and statistically significant, which is consistent with the logic that ad repetition leads to lower ad-tuning rates. We also find that the ad-creative fixed effects obtained from the specification with or without controlling for ad-wearout effects are highly correlated (.998). The second-stage results are qualitatively unchanged even after controlling for ads' cumulative viewership in the first stage.

Discussion

The research was motivated by the finding that the energy level in ads shown on TV has been increasing over time. Intrigued by this finding, we empirically investigate the relationship between the use of energetic ad content and the extent of ad-tuning time. A critical question is how energy is measured and what it represents. We estimate the energy level in ads by adapting a measure developed by Echo Nest, now owned by Spotify. To address what the measure represents, we validate the energy measure through an MTurk study. We find that the energy measure is strongly correlated with the construct of arousal. Arousal is directly linked to stimulation, which has been shown to affect ad tuning. The Spotify energy measure was estimated on music tracks, whereas our measure of energy in ads includes other background sounds, such as narration and product sounds, in addition to music. Further, visual aspects also contribute to the energy in an ad. We find through the MTurk study that the perceptions of audio energy and visual energy are strongly correlated, and that the Spotify energy measure is correlated to the perceived arousal of the audio part

of the ad, the perceived arousal of the visual part of the ad, and the combined audio and visual energy of the ad. To summarize, the empirically estimated Spotify energy measure is linked to arousal and to both the audio and visual energy of the ads.

We evaluate the relationship between the energy levels in TV commercials and ad-tuning rate using a large-scale data set, while controlling for network effects, program effects, an extensive set of time effects, and position effects of the ad creative. We find that, on average, longer ad-tuning rates are associated with higher energy in ad content. At the same time, we find that the result is not uniform and depends on the program genre and product categories. In the entertainment and news genres, we find a positive relationship between energy and ad-tuning rates in 11 of the 15 and 7 of the 15 product categories, respectively (Table 6; at the 10% significance level). A negative relationship is found in one product category in the news genre and three product categories in the sports genre.

Some findings in the literature, however, suggest that increases in energy may not always lead to increased viewing under certain conditions. Research by Puccinelli, Wilcox, and Grewal (2015), for instance, finds that highly energetic ads can lead to less watching when they are inserted in deactivating TV shows (e.g., sad movies). Cranton and Lantos (2011) state that "music that is either overly or insufficiently arousing for a particular consumer in a specific context will be regarded unfavorably" (p. 405) and recommend that the level of stimulation provided by the music be matched to that of the program or ad content. In a follow-up study, Lantos and Cranton (2012) point out that heterogeneity among viewers can lead to a positive or negative response to music in advertising, which can be extended to audio in general. These observations seem intuitive because judgments typically involve a reference to a standard, either internal or contextual. Thus, whether an ad is energetic might be evaluated in relation to the context. The implication is that the same level of arousal might be perceived as high in a low-arousal context but low in a high-arousal context. An important contextual factor in advertising therefore is the program in which the ad is run.

In real advertising settings, ads are seen in contexts with tremendous heterogeneity, which cannot be fully accounted for and controlled by advertisers. These observations may lead one to believe that the relationship between high energy in ads and the tendency to view ads longer (or shorter) will always depend on context; thus, that generalizable patterns are difficult to discern. Although there is truth to this view, we ask in our research what the relationship looks like after controlling for network, program, time, and ad position. Advertisers typically cannot choose the ad position within a program, and in some cases the specific program in which the ad is shown as well depending on how ad time is purchased. However, advertisers select the network and the year-week, day of the week, and hour of the day when their ads are shown. Our results suggest that, overall, there is a positive association between energy levels in ad content and one aspect of viewers' behavior, namely, ad tuning. We find that the

association, on average, is positive, but it varies both in magnitude and in direction across pairs of a program genre and a product category.

We also note that ad tuning may be affected by the presence and timing of when different objects appear in the ad. For instance, Teixeira, Wedel, and Pieters (2010) find that the timing of the appearance of the advertised brand's logo in an ad affects ad-tuning rates. Similarly, consumer reaction to ads can be affected by whether ads include animals (Trivedi and Teichert 2020). Our focus in this research, however, is on the role of an ad's energy in tuning behavior. Next, we discuss managerial implications of our findings.

Managerial Implications

Advertisers and television networks have routinely included audio in ads that is much louder than the audio in the programs in which they are aired. The assumption has been that an increase in the loudness of the audio of ads relative to that in the program attracts attention to the ads. Increasing attention to ads has been linked to lower ad avoidance (Teixeira, Wedel, and Pieters 2012; Tse and Lee 2001). The practice of making ads louder than the programs they are aired in became so prevalent that it raised concerns about the health effects of loudness on viewing audiences, leading to regulatory limitations on how much louder ads can be than the programs in which they are placed. The resulting CALM (Commercial Advertisement Loudness Mitigation) Act, which was passed in 2010, and began to be enforced by the Federal Communications Commission in 2012, limits the average loudness of an ad to no more than the average loudness of the program in which it is aired. Advertisers and networks, therefore, cannot continue to rely on loudness as a means of attracting attention to reduce ad avoidance. In addition, television manufacturers and streaming service providers have begun to offer features that give television audiences more control over the loudness of the programming. In principle, audiences can either mute ads or force them to be at the same level of sound as the programs they are aired in. Advertisers, therefore, need to be creative in how to use audio in their ads to attract and retain audience attention. This is a particularly vexing challenge because of the increasing problem of ads being avoided and skipped by viewers.

Our results suggest that, on average, increasing the energy in an ad can increase ad tuning or reduce ad avoidance. This finding is consistent with that of Belanche, Flavián, and Pérez-Rueda (2017). We also note that energy is not just loudness, and ad energy comprises both audio energy and visual energy. In our MTurk study, participants associated energetic ads with descriptors like "fast," "music," "movement," "upbeat," and "exciting." When testing different creatives, therefore, advertisers should also focus on evaluating the effects of different aspects of energy in the ability of ad creatives to gain higher viewing rates. Given our findings of heterogeneity, advertisers will also need to experiment and identify the program genres in which ad creatives with specific levels of energy are likely to be successful in attracting longer viewing rates.

Our empirical measure of energy includes all of the audio in the ad, which includes sounds associated with the visual aspects of the ad. A useful next step therefore would be to map levels of energy in ad content into a granular feature set, which includes both auditory components (e.g., dynamic range, entropy, loudness, onset rate, and timbre) and visual components (e.g., colorfulness, saturation, human facial expressions, motion, scene similarity). If a creative is being designed with background music, those composing the music should use all of these attributes to reach the target level of energy for the creative. For instance, they could increase the dynamic range of the music. Alternatively, they could increase the music's entropy, onset rate, or timbre to reach the target level. The same could be done with any speech or other visual elements in the creative. Then, the next step would be understanding how each component contributes to a certain aspect of consumer behavior, as well as how multiple components interact. This step would require data with (ideally exogenous) variation in each component. For instance, one could design an experiment in which the energy levels in audio and video of ads are randomized to tease out the relative role of the two variations, which we lack in our data.

Such strategic determination of ad content can benefit ad creators, as well as advertising publishers. Ad creators working with objective algorithms (e.g., MIR software) should be able to measure the values of these attributes for any auditory composition. They can then adjust the attributes to reach the targeted audio energy levels. Digital ad publishers, such as Spotify and Pandora, can also leverage the auditory attributes of tracks in users' listening sessions to better target ads. They could also inform advertisers to better design ad creatives based on the advertisers' target listeners.

Limitations and Directions for Future Research

The data used in this research are based on actual observations of the second-to-second tuning behaviors across millions of televisions rather than the smaller samples and specialized settings used in other studies of ad viewing behaviors. Several limitations, however, are in order. First, because of the observational nature of our data, our findings cannot be construed as capturing causal relationships. They are only correlational. Also, this study does not attempt to reveal the mechanism under which ad content affects consumer behavior. Second, we do not know whether the individual or individuals tuned in to the monitored televisions were indeed viewing or paying attention to the aired ads when the TV sets were tracked by our data provider. Relatedly, we have no data on other metrics that advertisers are interested in knowing, such as likability, memorability, recall, and conversion to purchase. Third, our data are missing consumer behaviors from time-shifted views, which are reported to be growing. Audience behaviors regarding ad avoidance may be different between live versus time-shifted views. Fourth, our data lack independent variation in the energy levels in ad audio and video, which prevents us from investigating any meaningful interaction between the two. Also, we focus

on the energy extracted from overall sound in ads rather than distinguishing between human voice and background sound. Lastly, our focus is on the overall association between energy in ad content and ad-tuning rate. As mentioned, however, in certain cases higher energy can lead to lower ad-tuning rates. Our results also show that the association varies across program genres and product categories, but we are unable to reveal the mechanism or mechanisms that explain when and how high energy in ad content is effective.

Although these limitations each point to a direction for future research on how the content of ads can influence their effectiveness, we believe an important next step is to further investigate our findings through controlled large-scale testing in a real-world context. We hope that this study motivates the initiation of such testing and provides initial guidelines for the design of such studies.

Associate Editor

Kenneth Wilbur

Authors' Note

The first two authors contributed equally to this work and are listed in reverse alphabetical order, and the last two authors are listed in alphabetical order. The authors are grateful to the editorial team for their excellent guidance. They have benefited from comments by Bobby Calder, Brett Gordon, Angela Lee, Gonca Soysal, Brian Sternthal, Ying Xie, Jie Zhang, and the seminar participants at the Kellogg Quant Workshop. The authors thank iSpot.tv for providing the data used in this research. The data provider is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All errors are those of the authors.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Lakshman Krishnamurthi would like to acknowledge the Montgomery Ward Foundation for their research support.

ORCID iD

Joonhyuk Yang  <https://orcid.org/0000-0002-2846-5868>

References

- Askin, Noah and Michael Mauskopf (2017), "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music," *American Sociological Review*, 82 (5), 910–44.
- Balducci, Bitty and Detelina Marinova (2018), "Unstructured Data in Marketing," *Journal of the Academy of Marketing Science*, 46 (4), 557–90.
- Belanche, Daniel, Carlos Flavián, and Alfredo Pérez-Rueda (2017), "Understanding Interactive Online Advertising: Congruence and Product Involvement in Highly and Lowly Arousing, Skippable Video Ads," *Journal of Interactive Marketing*, 37 (February), 75–88.
- Boughanmi, Khaled and Asim Ansari (2021), "Dynamics of Musical Success: A Machine Learning Approach for Multimedia Data Fusion," *Journal of Marketing Research*, 58 (6), 1034–57.
- Brost, Brian, Rishabh Mehrotra, and Tristan Jehan (2019), "The Music Streaming Sessions Dataset," in *WWW '19: The World Wide Web Conference*. New York: Association for Computing Machinery, 2594–2600.
- Cranton, Lincoln G. and Geoffrey P. Lantos (2011), "Attitude Toward the Advertising Music: An Overlooked Potential Pitfall in Commercials," *Journal of Consumer Marketing*, 28 (6), 396–411.
- Danaher, Peter J. (1995), "What Happens to Television Ratings During Commercial Breaks?" *Journal of Advertising Research*, 35 (1), 37–37.
- Defferrard, Michaël, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson (2017), "FMA: A Dataset for Music Analysis," in *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017, Suzhou, China, October 23-27, 2017*, Xiao Hu, Sally Jo Cunningham, Doug Turnbull, and Zhiyao Duan, eds. International Society for Music Information Retrieval, 316–23.
- Deng, Yiting and Carl F. Mela (2018), "TV Viewing and Advertising Targeting," *Journal of Marketing Research*, 55 (1), 99–118.
- Di Muro, Fabrizio and Kyle B. Murray (2012), "An Arousal Regulation Explanation of Mood Effects on Consumer Choice," *Journal of Consumer Research*, 39 (3), 574–84.
- Ferraro, Andrés, Dmitry Bogdanov, and Xavier Serra (2019), "Skip Prediction Using Boosting Trees Based on Acoustic Features of Tracks in Sessions," preprint, *arXiv*, 1903.11833, <https://arxiv.org/abs/1903.11833>.
- Fong, Hortense, Vineet Kumar, and K. Sudhir (2021), "A Theory-Based Interpretable Deep Learning Architecture for Music Emotion," working paper, Yale School of Management, <https://app.scholarsite.io/hortense-fong/articles/a-theory-based-interpretable-deep-learning-architecture-for-music-emotion>.
- Gorn, Gerald, Michel Tuan Pham, and Leo Yatming Sin (2001), "When Arousal Influences Ad Evaluation and Valence Does Not (and Vice Versa)," *Journal of Consumer Psychology*, 11 (1), 43–55.
- Jalali, Nima Y. and Purushottam Papatla (2016), "The Palette That Stands Out: Color Compositions of Online Curated Visual UGC That Attracts Higher Consumer Interaction," *Quantitative Marketing and Economics*, 14 (4), 353–84.
- Jehan, Tristan and Nicola Montecchio (2016), "Automatic Prediction of Acoustic Attributes from an Audio Signal," U.S. patent US10089578B2, <https://patents.google.com/patent/US10089578B2/en>.
- Lantos, Geoffrey P. and Lincoln G. Cranton (2012), "A Model of Consumer Response to Advertising Music," *Journal of Consumer Marketing*, 29 (1), 22–42.
- Lewis, Jeffrey B. and Drew A. Linzer (2005), "Estimating Regression Models in Which the Dependent Variable Is Based on Estimates," *Political Analysis*, 13 (4), 345–64.
- Li, Xi, Mengze Shi, and Xin Shane Wang (2019), "Video Mining: Measuring Visual Information Using Automatic Methods," *International Journal of Research in Marketing*, 36 (2), 216–31.
- Liu, Liu, Daria Dzyabura, and Natalie Mizik (2020), "Visual Listening In: Extracting Brand Image Portrayed on Social Media," *Marketing Science*, 39 (4), 669–86.

- Liu, Xiao, Param Vir Singh, and Kannan Srinivasan (2016), "A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing," *Marketing Science*, 35 (3), 363–88.
- Lu, Shasha, Li Xiao, and Min Ding (2016), "A Video-Based Automated Recommender (VAR) System for Garments," *Marketing Science*, 35 (3), 484–510.
- Marinova, Detelina, Sunil K. Singh, and Jagdip Singh (2018), "Frontline Problem-Solving Interactions: A Dynamic Analysis of Verbal and Nonverbal Cues," *Journal of Marketing Research*, 55 (2), 178–92.
- McFee, Brian, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, et al. (2015), "librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference*, Austin, TX, http://conference.scipy.org/proceedings/scipy2015/brian_mcfree.html.
- McGranaghan, Matthew, Jura Liaukonyte, and Kenneth C. Wilbur (2021), "TV Ad Viewability: How Viewer Tuning, Presence and Attention Respond to Ad Content," working paper, <https://kennethcwilbur.github.io/website/MLW2021%20TV%20Ad%20Viewability.pdf>.
- Mehrabian, Albert and James A. Russell (1974), *An Approach to Environmental Psychology*. Cambridge, MA: MIT Press.
- Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31 (3), 521–43.
- Ng, Aaron and Rishabh Mehrotra (2020), "Investigating the Impact of Audio States & Transitions for Track Sequencing in Music Streaming Sessions," in *RecSys '20: Fourteenth ACM Conference on Recommender Systems*. New York: Association for Computing Machinery, 697–702.
- Olney, Thomas J., Morris B. Holbrook, and Rajeev Batra (1991), "Consumer Responses to Advertising: The Effects of Ad Content, Emotions, and Attitude Toward the Ad on Viewing Time," *Journal of Consumer Research*, 17 (4), 440–53.
- Park, Minsu, Jennifer Thom, Sarah Mennicken, Henriette Cramer, and Michael Macy (2019), "Global Music Streaming Data Reveal Diurnal and Seasonal Patterns of Affective Preference," *Nature Human Behaviour*, 3 (3), 230–36.
- Preston, Ivan L. (1982), "The Association Model of the Advertising Communication Process," *Journal of Advertising*, 11 (2), 3–15.
- Puccinelli, Nancy M., Keith Wilcox, and Dhruv Grewal (2015), "Consumers' Response to Commercials: When the Energy Level in the Commercial Conflicts with the Media Context," *Journal of Marketing*, 79 (2), 1–18.
- Russell, James A. (1980), "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, 39 (6), 1161–78.
- Siddarth, Sivaramakrishnan and Amitava Chattopadhyay (1998), "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17 (2), 124–38.
- Teixeira, Thales, Rosalind Picard, and Rana El Kaliouby (2014), "Why, When, and How Much to Entertain Consumers in Advertisements? A Web-Based Facial Tracking Field Study," *Marketing Science*, 33 (6), 809–27.
- Teixeira, Thales, Michel Wedel, and Rik Pieters (2010), "Moment-to-Moment Optimal Branding in TV Commercials: Preventing Avoidance by Pulsing," *Marketing Science*, 29 (5), 783–804.
- Teixeira, Thales, Michel Wedel, and Rik Pieters (2012), "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 49 (2), 144–59.
- Trivedi, Rohit H. and Thorsten Teichert (2020), "Consumer Reactions to Animal and Human Models in Print Ads: How Animals and People in Ads Influence the Purchase-Decision Journey," *Journal of Advertising Research*, 60 (4), 426–38.
- Tse, Alan Ching Biu and Ruby P.W. Lee (2001), "Zapping Behavior During Commercial Breaks," *Journal of Advertising Research*, 41 (3), 25–29.
- Tuchman, Anna E., Harikesh S. Nair, and Pedro M. Gardete (2018), "Television Ad-Skipping, Consumption Complementarities and the Consumer Demand for Advertising," *Quantitative Marketing and Economics*, 16 (2), 111–74.
- Van Meurs, Lex (1998), "Zapp! A Study on Switching Behavior During Commercial Breaks," *Journal of Advertising Research*, 38 (1), 43–44.
- Wilbur, Kenneth C. (2008), "How the Digital Video Recorder (DVR) Changes Traditional Television Advertising," *Journal of Advertising*, 37 (1), 143–49.
- Wilbur, Kenneth C. (2016), "Advertising Content and Television Advertising Avoidance," *Journal of Media Economics*, 29 (2), 51–72.
- Woltman Elpers, Josephine L.C.M., Michel Wedel, and Rik G.M. Pieters (2003), "Why Do Consumers Stop Viewing Television Commercials? Two Experiments on the Influence of Moment-to-Moment Entertainment and Information Value," *Journal of Marketing Research*, 40 (4), 437–53.
- Xiao, Li and Min Ding (2014), "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science*, 33 (3), 338–52.
- Xiao, Li, Hye-jin Kim, and Min Ding (2013), "An Introduction to Audio and Visual Research and Applications in Marketing," *Review of Marketing Research*, 10, 213–53.