Learning From Highly Unstructured Data: Insight From Videos, Images and Audio

Reto Hofstetter, University of Lucerne
Jonah Berger, penn
Aradhna Krishna, Univ. of Michigan, Ann Arbor
Marc Bravin, University of Applied Sciences and Arts Lucerne
Melanie Clegg, University of Lucerne
Marc Pouly, University of Applied Sciences and Arts Lucerne
Prashant Rajaram, Ivey Business School at Western University
Puneet Manchanda, University of Michi
Jochen Hartmann, University of Hamburg
Verena Schoenmueller, Bocconi University
Yonat Zwebner, IDC Herzliya
Jacob Goldenberg, Interdisciplinary Center
Oded Netzer, Columbia Business School
Na Kyong Hyun, Georgia Tech
Michael Lowe, Georgia Tech

Consumer researchers and marketing managers alike are increasingly confronted with highly unstructured data. Text, images, audio, and video provide valuable sources of information, but this content is often non-numeric, multi-faceted, and difficult to parse. How can we better use these exciting resources to test theory and uncover insight?

[to cite]:

Reto Hofstetter, Jonah Berger, Aradhna Krishna, Marc Bravin, Melanie Clegg, Marc Pouly, Prashant Rajaram, Puneet Manchanda, Jochen Hartmann, Verena Schoenmueller, Yonat Zwebner, Jacob Goldenberg, Oded Netzer, Na Kyong Hyun, and Michael Lowe (2021) ,"Learning From Highly Unstructured Data: Insight From Videos, Images and Audio", in NA - Advances in Consumer Research Volume 49, eds. Tonya Williams Bradford, Anat Keinan, and Matthew Matthew Thomson, Duluth, MN : Association for Consumer Research, Pages: 808-813.

[url]:

http://www.acrwebsite.org/volumes/3000904/volumes/v49/NA-49

Learning from Highly Unstructured Data: Insight from Videos, Images and Audio

Chairs: Reto Hofstetter, University of Lucerne, Switzerland
Jonah A. Berger, The Wharton School, USA
Discussant: Aradhna Krishna, University of Michigan, USA

Paper #1: Does it help to be creative on TikTok?

Marc Bravin, University of Applied Sciences and Arts Lucerne, Switzerland
Melanie Clegg, University of Lucerne, Switzerland
Reto Hofstetter, University of Lucerne, Switzerland
Marc Pouly, University of Applied Sciences and Arts Lucerne, Switzerland
Jonah A. Berger, The Wharton School, USA

Paper #2: Video Influencers: Unboxing the Mystique

Prashant Rajaram, Ivey Business School at Western University, USA
Puneet Manchanda, University of Michigan, USA

Paper #3: Brand Faces: Mining Brand Preferences from Consumer Faces

Jochen Hartmann, University of Hamburg, Germany
Verena Schönmüller, Bocconi University, Italy
Yonat Zwebner, IDC Herzliya, Israel
Jacob Goldenberg, IDC Herzliya, Israel
Oded Netzer, Columbia University, USA

Paper #4: Measuring Objective Vocal Similarity in Human-AI Agent Interactions

Na Kyong Hyun, Georgia Tech, USA
Michael Lowe, Georgia Tech, USA
Aradhna Krishna, University of Michigan, USA

SESSION OVERVIEW

Consumer researchers and marketing managers alike are increasingly confronted with highly unstructured data. Text, images, audio, and video provide valuable sources of information, but this content is often non-numeric, multi-faceted, and difficult to parse (Balducci & Marinova, 2018). How can we better use these exciting resources to test theory and uncover insight?

While some prior work has begun to explore unstructured data, the applicability of these methods in consumer research is still quite novel. How can we use video content to understand what becomes popular and why? Might profile pictures provide insight into personality and thus the brands people prefer? And how do vocal features impact perceptions of warmth and competence? This session addresses these and other questions as it deepens our understanding of these new sources of data, and how researcher can parse them to better understand consumer behavior.

First, Bravin, Clegg, Hofstetter, Pouly, and Berger employ a 3D convolutional neural network trained on a large data set of TikTok videos to investigate the value of originality in social media content. Analysis of hundreds of thousands of posts, combined with controlled experiments reveals that highly original content can backfire, because it evokes cognitive dissonance.

Second, Rajaram and Manchanda use deep learning to analyze unstructured data across multiple modalities (text, audio and images). They provide a novel *interpretable* methodological approach that allows testing causality and theory in various contexts. They demonstrate their approach through examining how the composition of influencers' video content can influence user reactions.

Third, Hartmann, Schönmüller, Zwebner, Goldenberg, and Netzer apply deep learning to automated face analysis, testing relationships between personality factors extracted from profile pictures and consumer brand preferences. Their findings validate a conceptual link between consumer facial features and brand personality dimensions and imply how image analysis can be used to predict consumer behavior.

Fourth, Na Kyong, Lowe, and Krishna measure the timbre similarity between users' and AI voices. Three studies analyze the influence of vocal similarity on perceived warmth and competence of AI agents, and choice. They demonstrate how automated voice analyses can shed light on the impact of subtle perceptual cues.

Given the dramatic growth of unstructured data, the world needs a better understanding of how to use it to build consumer theory. The papers in this special session provide methodological tools that offer innovative and theory-driven solutions for important issues in consumer research. These tools are widely applicable to problems related to high-scale and unstructured data sources, as is the case in many of today's contexts (e.g., social media, online branding, and online behavior). Furthermore, this session motivates researchers to embrace methodological triangulation to solve current consumer research issues.

Does it Help to Be Creative on TikTok?

EXTENDED ABSTRACT

People often suggest that originality should boost success on social media. Particularly in online contexts, people like to talk about and share novel information (Berger & Iyengar, 2013) as it is considered more surprising, entertaining, and useful resulting in greater proliferation (Berger, 2014; Berger & Milkman, 2012). Since creative and original things appeal to our natural curiosity (Silvia, 2008), it is widely assumed that originality is valuable in social and business contexts (Hofstetter, Dahl, Aryobsei, & Herrmann, 2020).

But is that actually true? Psychological theories suggest that less originality may actually increase liking of stimuli in our environment. For instance, seeing the same stimulus several times positively affects our attitudes toward this stimulus—a phenomenon called "mere exposure effect" (Zajonc, 1960). Conversely, there exists a systematic bias against high novelty, as it harms our expectations and questions our existing cognitive patterns (Eidelman, Crandall, & Pattershall, 2009; Toubia & Netzer, 2017). Indeed, on social media platforms, content that is of surprisingly low originality oftentimes becomes viral: On TikTok, trending videos follow the same dancing and lip sync screenplay (e.g., one million for the song "savage love", Aniftos 2020).

Accordingly, we suggest that high originality is indeed not advantageous for social media content. Based on cognitive dissonance theory, we argue that highly original content can trigger uncertainty and psychological discomfort (i.e., cognitive dissonance) among viewers, which is an unpleasant state (Whitson & Galinsky 2008). Three studies provide support for this expectation.

The first study provides empirical evidence for a negative effect of originality on content liking. We collected and analyzed a large dataset of 290,503 videos from the social media platform TikTok. The dataset contains descriptive video information (e.g., hashtags, music played, etc.) and content-related statistics such as the number of views, likes, shares, and comments. Since no labels indicate video dissimilarities, we implement a self-supervised contrastive learning model inspired by Qian et al. (2020). The model is a 3D convolutional neural network trained by simultaneously maximizing the similarity between transformed views of the same video and minimizing the similarity between transformed views of different videos. It allows to extract semantic video embeddings that were used for further analysis. To estimate the degree of originality of a video, we employ the local outlier factor (LOF) algorithm (Breunig, Kriegel, Ng, & Sander, 2000), which computes the local density deviation of a given data point with respect to its k neighbors. Less original videos are more likely to end up in a cluster with high density and thus have a lower LOF. We then use this originality measure in four statistical models that vary in terms of the included control variables to explain video liking. In our first model, we find a significant negative effect of originality on the number of likes ($\beta_{Originality}$=-1.07, p<.001). Hence, a 1% increase in originality leads to about a 1% reduction in likes. We control for user and music fixed effects in models 2 and 3, respectively, and find that the coefficient decreases consistently with this assumption. Model 4 includes four additional controls (number of days the video was online, popularity, and originality × popularity). We find that a video's lifespan increases its likes ($\beta_{Nr\ of\ days\ online}$=.19, p<.001) and that the later a user adopted a song in the sequence of adopters, the more likes the video receives (interpreted as higher popularity, $\beta_{Popularity\ (Nth\ song\ user)}$=.33, p<.001). We find a significant interaction between popularity and originality ($\beta_{Popularity\ (Nth\ user\ of\ song)}$=.04, p<.001), showing that the harmful influence of originality is dampened for later adopters.

These models suggest that more original videos receive fewer likes. The fact that the effect is reduced for later adopters supports a cognitive dissonance explanation. As a screenplay becomes more popular, greater originality is required to stand out and trigger dissonance.

Studies two and three test causality through manipulating content originality and measuring liking. We keep the focal content the same but manipulate the context. Everyone saw the same target video (a particular dance), and rated how much they liked it, but we varied the similarity of some videos they watched beforehand. In the low originality condition, participants watched three videos of the same dance first, while in the high originality condition they saw three videos from the same users that involved a completely different dance. All videos were originally retrieved from TikTok. Participants indicated their likelihood to "like" (i.e., how likely they would click on the like button) the video on a scale from 1 (very unlikely) to 7 (very likely).

In study 2 (N=192 participants recruited via Prolific), we find support for the negative main effect of content originality. One-way ANOVA shows that participants are less likely to "like" the target video if it is more original and less similar to previously seen content ($M_{Low\ originality}$=4.17, SD=2.05, $M_{High\ originality}$=3.55, SD= .18, F(1, 194)=4.04, p=.046, d=-.29).

In study 3 (N=134, MTurk), we used different videos, and included a scale for cognitive dissonance (Jiang, Hoegg, & Dahl, 2013). We find again a negative main effect of originality on liking ($M_{Low\ originality}$=4.36, SD=2.14, $M_{High\ originality}$=3.42, SD=2.36, F(1, 132)=7.58, p=.017, d=-.42), which is mediated by cognitive dissonance (indirect effect=-.26, SE=.13, $CI_{95\%}$=[-.04,-.55]; analysis based on 10,000 bootstrapped resamples, 95% confidence intervals; Hayes, 2013).

Our findings support our conceptual reasoning that higher originality of content on social media does not generally pay out. These findings are in contrast to prior research promoting the high value of originality of user-generated content (Berger & Milkman, 2012; Hofstetter et al., 2020) and suggest a potential negative mechanism triggered by increased cognitive dissonance when viewing highly original content. Our research contributes to literature on consumer behavior on social media and advises managers to not overestimate the value of highly original content (e.g., for brand-related user-generated content). A methodological contribution of our research comes from the development of a self-supervised learning approach that allows systematic investigations of video-based social media content by quantifying its originality.

Video Influencers: Unboxing the Mystique

EXTENDED ABSTRACT

Influencers have the capacity to shape the opinion of others in their network. They were traditionally celebrities (e.g., movie stars and athletes) who leveraged their expertise, fame and following in their activity domain to other domains. However, 95% of the influencers today, or "social media stars," are individuals who have cultivated an audience over time by making professional content that demonstrates authority and credibility (Creusy, 2016; O'Connor, 2017). The growth in their audience(s) has been in part attributed to the fact that influencer videos are seen as "authentic" based on a perception of high source credibility. The increasing popularity of social media stars has resulted in an exponential growth of the influencer marketing industry which is expected to reach a global valuation of $15B in 2022 from $8B in 2019 (Business Insider, 2021). There are now more than 1100 influencer marketing agencies in the world that allow brands to partner with influencers to promote their products (Influencer Marketing Hub and CreatorIQ, 2020). These influencers primarily reach their audience(s) via custom videos that are available on a variety of social media platforms (e.g., YouTube, Instagram, Twitter and TikTok) (Brooks, 2020). Despite the rapid emergence and growth of influencer videos, there is limited research on their design and effectiveness. Specifically, little is known about the relationship between video content and viewer reactions as well as the evolution of these videos over time.

In this paper, we investigate whether the presence and nature of advertising content in videos is associated with relevant outcomes (views, interaction rates, and sentiment). There are a few challenges in carrying out these tasks. First, most data in influencer videos are unstructured. In addition, these data span different modalities – text, audio and images. This necessitates the use of state-of-the-art machine learning methods commonly referred to as deep learning. The second challenge arises from the fact that past approaches in marketing using such methods have typically made a tradeoff between predictive ability and interpretability. Specifically, such deep learning models traditionally use unstructured data to predict marketing outcomes well out-of-sample but suffer from poor interpretability. On the other hand, deep learning models that use ex-ante handcrafted features obtain high interpretability of the captured relationships but suffer from poor predictive ability. Our "interpretable deep learning" approach uses unstructured data across multiple modalities (text, audio and images) to make predictions out-of-sample and ex-post interprets the machine learning "black-box", thus avoiding the need to make this trade-off. We apply our approach to a random sample of publicly available videos of 33 YouTube influencers who receive brand sponsorship.

Our approach helps us identify statistically significant relationships between marketing (brand) relevant outcomes and video elements. The significance of these relationships is supported by a significant change in attention (importance) paid by the model to these video elements. For the outcomes, we use publicly available data to develop metrics based on industry practice (Influencer Marketing Hub and CreatorIQ, 2020) and past research on visual and verbal components of conventional advertising (Mitchell, 1986). These metrics are # views, engagement (#comments / # views), popularity (# likes / # views), likeability (# likes / # dislikes) and sentiment. The influencer video elements we consider are text (e.g., brand names in title, captions/transcript and description), audio (e.g., speech, music, etc.), and images (e.g., brand logos, persons, clothes, etc. in thumbnails and video frames). In the ex-post interpretation step, we identify salient word pieces in text, moments in audio and pixels in images.

The focus on interpretation allows us to document some interesting relationships (based on a holdout sample) across all three modalities. First, we find that brand name inclusion, especially in the consumer electronics and video game categories, in the first 30 seconds of captions/transcript is associated with a *significant increase* in attention paid to the brand but a *significant decrease* in predicted sentiment. Second, music (without simultaneous human sound) within the first 30 seconds is associated with a *significant increase* in attention. However, longer music duration is associated with a *significant decrease* in predicted engagement, popularity and likeability but a *significant increase* in predicted sentiment. Third, larger pictures (of persons as well as clothes & accessories) in five equally spaced video frames (within the first 30 seconds) are associated with a *significant increase* in attention and predicted engagement. Finally, we also demonstrate that the focus on interpretability does not compromise the predictive ability of our model.

These results are relevant for multiple audiences. For academics, who may be interested in testing causal effects, our approach is able to identify a smaller subset of relationships for formal causal testing. This is done by filtering out more than 50% of relationships that are affected by confounding factors unassociated with attention (importance) paid to video elements. For practitioners, we provide a general approach to the analysis of videos used in marketing that does not rely on primary data collection. For brands, influencers and agencies, our results provide an understanding of the association between video features and relevant outcomes. Influencers can iteratively refine their videos using our model and results to improve performance on an outcome of interest. Brands, on the other hand, can evaluate influencer videos to determine their impact and effectiveness at various levels of granularity (individual video elements, interactions of elements or holistic influence).

Overall, this paper makes four main contributions. First, to the best of our knowledge, it is the first paper that rigorously documents the association between advertising content in influencer videos and marketing outcomes. Second, it presents an interpretable deep learning approach that avoids making a tradeoff between interpretability and predictive ability. It not only predicts well out-of-sample but also allows interpretation and visualization of salient regions in videos across multiple data modalities – text, audio, and images. Third, it generates novel Hypothesis between advertising content and a change in the outcome of interest for formal causal testing as noted above. Finally, it provides a comprehensive, data-based approach for marketers (and influencers) to assess and evaluate the quality of videos.

Brand Faces: Mining Brand Preferences from Consumer Faces

EXTENDED ABSTRACT

For centuries, philosophers, researchers, and practitioners have been fascinated by the information contained in faces and conveyed by facial expressions. This research investigates the association between brand preferences and faces ("brand faces"). Specifically, we propose a novel multi-method approach to extract brand preferences from consumers' profile pictures on social media. While faces have already been used to successfully predict names (Zwebner et al. 2017), sexual orientation (Wang & Kosinski 2018), and political affiliation (Tkachenko & Jedidi 2019), predicting brand preferences from consumers' digital self-portraits poses a promising and managerially relevant challenge.

Conveying identity and emotion, faces are a cornerstone of human communication. Our research bridges two recent literature streams. First, it links to the automated face analysis literature in marketing, management, and psychology (e.g., Choudhury et al. 2019; McDuff & Berger 2020, Xiao & Ding 2014, Zwebner et al. 2017). Second, it employs automated social media mining to harvest brand-related information (e.g., Culotta & Cutler 2016; Netzer et al. 2012). Building on extant research establishing a link between identities and brands (e.g., Bellazza & Berger 2019; Berger & Heath 2007) as well as between identities and faces (e.g., Ballew & Todorov 2007; Cogsdill, et al., 2014; Kachur et al. 2020), we explore if a direct link between faces and brands also manifests in real-world social media data.

Employing recent deep learning techniques for automated face analysis, we show what brand-related information is conveyed in the face by the way consumers present themselves on social media. Specifically, we obtained a data set containing more than 100,000 single-face profile pictures and each user's followership across 444 brands from more than 20 categories (e.g., apparel, cars, print media). For each consumer we represent his or her face as a low-dimensional embedding with the objective to encode it as efficiently as possible while retaining its distinctive characteristics. For this purpose, we work with both a commercial solution (MS Azure) and validate results with a specialized deep neural network (DNN) pre-trained on 2.6 million faces (i.e., the VGG-Face architecture by Oxford's Visual Geometry Group, Parkhi et al. 2015). The DNN embeds faces as 2,622-dimensional vectors through a sequence of (non-)linear operations (see Wang & Kosinski (2018) for a similar approach). The commercial solution returns predictions for 27 facial features, i.e., demographics (e.g., age, gender), head position (pitch, roll, yaw), emotions (e.g., happiness, sadness, anger), accessories (e.g., sunglasses), image quality (e.g., blur, noise).

Across three complementary analyses, we demonstrate the relationship between consumer faces and brand preferences. First, for more than 7,500 unique within-category brand pairs, including "canonical competitors" such as Apple vs. Microsoft, we predict consumers' brand preferences just from their faces significantly above the random-chance baseline. For all brand pairs users with overlapping brand preferences, i.e., users following both brands, are excluded. Out of these images a balanced sample of 200 images (100 per brand) is drawn (80% training data, 20% test data). Given the limited training data, the accuracy levels can be considered conservative estimates. Specifically, the global distribution of accuracies reveals a mean accuracy level of nearly 60%. Hence, error rates are reduced by about 20% compared to a random-chance baseline. Results from robustness checks that we run on subsets of the data suggest that our models capture predictive signals from consumers' profile pictures beyond socio-demographic information. Second, zooming in on consumer faces, our data reveal insightful associations between 27 specific facial features and 47 established brand personality dimensions from the Young & Rubicam's Brand Asset Valuator (BAV), e.g., wearing reading glasses is predictive of liking "intelligent" brands while wearing lip and eye makeup predict liking "glamorous" brands. Third, we can recover meaningful market maps across brands solely from consumer faces, which we validate with external brand perception data. To preempt possible privacy concerns (see Van Noorden (2020) and Zhou et al. (2020) for recent discussions), we abstract from individual faces across all our analyses and explore the link between consumer faces and brand preferences only from an aggregated perspective.

Overall, our results highlight the power of automated face-based brand audience analytics and reveal important implications not only for marketers but also for policy makers as well as for consumers themselves. There are, of course, limitations to our approach. First, brand followership is only a proxy for brand affiliation. For example, someone may follow brands (or politicians) rather out of curiosity and entertainment purposes rather than honest brand preference (see Schoenmueller et al. 2021 for a similar approach). Second, consumers may disguise their real face on social networks. Their profile pictures are not a representation of who they are, but instead of how they consciously choose to present themselves and want to be perceived by others. Despite these limitations, we hope our results stimulate future research on the opportunities and challenges of automated face analytics.

Measuring Objective Vocal Similarity in Human-AI Agent Interactions

EXTENDED ABSTRACT

AI agents (e.g., Apple Siri, Amazon Alexa, social robots including Pepper) have become increasingly prevalent in our daily lives, our homes, and into our workplaces. Voice recognition technology via machine learning has reached the accuracy level of human speech (Arnold, 2018) and algorithms will eventually recognize all the various aspects in speech including nuances and vocal characteristics (i.e., tonal inflection, mood; Kirby, 2019). Also, the use of voice recognition will be advanced to include personalization features, similar to face ID or PIN, so that the AI agent can respond accordingly to the identified user through the vocal characteristics unique to each person's vocal tract. This research examines vocal similarity between an individual consumer and the AI agent. We ask this question: how will consumers perceive, respond to, and be persuaded by an AI agent contingent on similarities between that agent's voice and their own voice?

In this research, we focus on timbre since it is the most prominent feature allowing humans or machines to distinguish one voice from another. Timbre refers to the unique spectrum of frequencies *within* a sound, including one's voice (here measured by Mel Frequency Cepstral Coefficients [MFCCs; Logan, 2000]) and largely accounts for our ability to distinguish between voices or instruments, and can also affect consumer perception in various ways (Bruner 1990). In this work, we ask how *differences* in timbre between a consumer's voice and an AI voice might influence diverse consumer attitudes toward the AI agent, and further impact consumption choice.

This objective measure of vocal similarity in timbre (MFCCs) introduces a new methodological approach measuring the Euclidian distance between the AI agent and each participant's voice. In order to objectively capture and quantify timbre similarity, we wrote employ an observed machine learning process using MATLAB that enables the machine to learn and calculate a point allowing the total Euclidean distance of all vocal elements to reach a minimum. Then, the similarity points are calculated by using the Euclidean distance between the lower rank matrices of the compared voices (i.e., VocSimilarity = dist(Female1, AI Voice1')). Using this procedure to calculate speech similarity, we test our proposed research question and Hypothesis in three experimental studies.

A prevalent body of research in social psychology advocates for the similarity-attraction effect (Collisson & Howell, 2014; Montoya, Horton, & Kirchner, 2008), which suggests that we are more likely to prefer and be initially attracted to similar others. Although research suggests that we prefer to interact with advanced machines that are more human in nature (Nass et al., 1995; Tapus & Mataric, 2007), the uncanny valley theory (Mori, 1970) suggests that the degree to which these machines are similar may play a role in that we experience an eerie sensation and discomfort when they become too similar to ourselves.

In our series of studies, after listening to a single (study 1 and 2) or three different (study 3) AI agents' recommendation of different products (books, SNL videos, and movies), we measured participant's perception of the AI agent. Then, we later recorded and analyzed each participant's voice to create an objective measure of acoustic similarity to the AI voice.

In study 1, participants (152 undergraduates) listened to a sample of an AI voice (gender matched) making a book recommendation and provided their impressions regarding the AI agent. Each participant also had a sample of their voice recorded upon conclusion of the study. Objective similarity between AI and participant voices was calculated using a computer algorithm created for this research. Overall, greater similarity in timbre (MFCCs) led to perceptions of greater warmth (MFCC: $B = -.402$, $SE = .199$, $p < .05$) and competence (MFCC: $B = -.396$, $SE = .159$, $p < .05$). Study 2 (187 undergraduates), used a similar procedure to Study 1, while also including a choice task based on the AI agent's video recommendation. The results suggested that the more similar the participant's voice was to the AI agent's voice in terms of timbre (MFCCs), the participant was more likely to choose the video recommended by the AI agent (MFCC: $B = -.202$, $SE = .087$, $p < .05$).

In Study 3, we introduce three distinct voices of AI agents recommending three different movies. In this study, we demonstrate the effect of vocal similarity on persuasion and trust (competence, benevolence, integrity), and further show that trust mediates actual choice in the recommended movie. The results again show that similarity in timbre significantly predicts evaluations of warmth (MFCC: $B = -.323$, $SE = .085$, $p < .001$), competence (MFCC: $B = -.289$, $SE = .082$, $p < .001$), liking (MFCC: $B = -.295$, $SE = .082$, $p < .001$), and overall trust in the agent (MFCC: $B = -.155$, $SE = .068$, $p < .05$). We also tested the proposed mediation model using the PROCESS macro model 4 (Hayes, 2013). The data shows a full mediation model (direct effect: b = -.231, $SE = .146$, $p > .05$, 95% CI [-518, .055]; indirect effect: b = -.093, $SE = .046$ 95% CI [-.191, -.009]) where overall timbre similarity impacts trust (b = -.155, $SE = .067$, $p < .05$), which in turn leads to higher likelihood to choose the recommended movie (b = .606, $SE = .102$ $p < .001$).

Present research aims to contribute to literatures in psychoacoustics, similarity-attraction effect, and human-computer interaction. We find that, overall, similar timbre and dissimilar pitch was favored. Furthermore, we believe that our research offers useful implications to marketers regarding new technology devices. Although certain voice-relevant cues (e.g., accent, conversational styles) may be more direct and pronounced to individuals when engaging in an interaction, we explore a very subtle cue that subconsciously influence consumers' mindset. Marketers will be able to better understand the mechanisms and conditions under which we prefer AI agents that are more personalized to sound similar in timbre to individual consumers.

REFERENCES

Aniftos, R. (2020). The 25 Best Musical TikTok Trends of 2020. Retrieved from https://www.billboard.com/articles/news/list/9492366/best-musical-tiktok-trends-2020-top-25/.

Arnold (2018). *A.I. and speech advances bring virtual assistants to work.* Retrieved from https://www.computerworld.com/article/3264433/ai-and-speech-advances-bring-virtual-assistants-to-work.html.

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557-590.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, 104(46), 17948-17953.

Bellezza, S., & Berger, J. (2020). Trickle-round signals: When low status is mixed with high. *Journal of Consumer Research*, 47(1), 100-127.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology, 24*(4), 586-607.

Berger, J., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, 34(2), 121-134.

Berger, J., & Iyengar, R. (2013). Communication Channels and Word of Mouth: How the Medium Shapes the Message. *Journal of Consumer Research, 40*(3), 567-579.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192-205.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). *LOF: identifying density-based local outliers*. Paper presented at the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, Texas, USA.

Brooks, A. (2020, January 24). As influencers increasingly create video content, what does this mean for brands? Retrieved from https://marketingtechnews.net/news/2020/jan/24/influencers-increasingly-create-video-content-what-does-mean-brands/.

Bruner, Gordon C. (1990), "Music, Mood, and Marketing," *Journal of Marketing* 54, (4), 94-104.

Business Insider. (2021). Influencer Marketing: Social media influencer market stats and research for 2021. Retrieved from https://www.businessinsider.com/influencer-marketing-report#:~:text=The%20influencer%20marketing%20industry%20is,gold%20standard%20for%20the%20group.

Chattopadhyay, Amitava, Darren W. Dahl, Robin JB Ritchie, and Kimary N. Shahin (2003), "Hearing Voices: The Impact of Announcer Speech Characteristics on Consumer Response to Broadcast Advertising," *Journal of Consumer Psychology,* 13 (3), 198-204.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, 40(11), 1705-1732.

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, 25(5), 1132-1139.

Creusy, K. (2016, July 12). What is influencer marketing? Retrieved from https://www.upfluence.com/influencer-marketing/what-is-influencer-marketing.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing Science*, 35(3), 343-362.

Eidelman, S., Crandall, C. S., & Pattershall, J. (2009). The Existence Bias. *Journal of Personality and Social Psychology, 97*(5), 765-775.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press.

Hofstetter, R., Dahl, D. W., Aryobsei, S., & Herrmann, A. (2020). Constraining Ideas: How Seeing Ideas of Others Harms Creativity in Open Innovation. *Journal of Marketing Research*, 0022243720964429.

Influencer Marketing Hub and CreatorIQ. (2020). *The State of Influencer Marketing 2020 : Benchmark Report* Retrieved from https://influencermarketinghub.com/influencer-marketing-benchmark-report-2020/.

Jiang, L., Hoegg, J., & Dahl, D. W. (2013). Consumer reaction to unearned preferential treatment. *Journal of Consumer Research, 40*(3), 412-427.

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. *Scientific Reports*, 10(1), 1-11.

Kirby (2019). *What's The Future of Voice Recognition?* Retrieved from https://www.futureofeverything.io/future-voice-recognition/.

Krishna, Aradhna and Norbert Schwarz (2014), "Sensory Marketing, Embodiment, and Grounded Cognition: A Review and Introduction." *Journal of Consumer Psychology,* 24, (2), 159-168.

Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling," In *Ismir*, vol. 270, pp. 1-11. 2000.

McDuff, D., & Berger, J. (2020). Why do some advertisements get shared more than others?: Quantifying facial expressions to gain new insights. *Journal of Advertising Research*, 60(4), 370-380.

Mitchell, A. A. (1986). The effect of verbal and visual components of advertisements on brand attitudes and attitude toward the advertisement. *Journal of consumer research, 13*(1), 12-24.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889-922.

Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33-35.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities?. *International Journal of Human-Computer Studies*, *43*(2), 223-239.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.

O'Connor, C. (2017, September 26). Forbes Top Influencers: Meet The 30 Social Media Stars Of Fashion, Parenting And Pets (Yes, Pets). Retrieved from https://www.forbes.com/sites/clareoconnor/2017/09/26/forbes-top-influencers-fashion-pets-parenting/#1a67cea27683.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

Qian, R., Meng, T., Gong, B., Yang, M. H., Wang, H., Belongie, S., & Cui, Y. (2020). Spatiotemporal contrastive video representation learning. *arXiv preprint arXiv:2008.03800.*

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.

Schoenmueller, V., Netzer, O., & Stahl, F. (2021). Polarized America: From political partisanship to preference partisanship. Working Paper.

Silvia, P. J. (2008). Interest—The Curious Emotion. *Current Directions in Psychological Science,* 17(1), 57-60.

Tapus, A., & Mataric, M. J. (2007, March). Emulating Empathy in Socially Assistive Robotics. In *AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics* (pp. 93-96).

Tkachenko, Y., & Jedidi, K. (2019). What's in a face? Predicting consumer information from facial images. Working Paper.

Toubia, O., & Netzer, O. (2017). Idea Generation, Creativity, and Prototypicality. *Marketing Science, 36*(1), 1-20.

Van Noorden, R. (2020). The ethical questions that haunt facial-recognition research. *Nature*, 587(7834), 354-358.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 114(2), 246.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, *322*(5898), 115-117.

Xiao, L., & Ding, M. (2014). Just the faces: Exploring the effects of facial features in print advertising. *Marketing Science*, 33(3), 338-352.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2), 1-27.

Zhou, Y., Lu, S., & Ding, M. (2020). Contour-as-Face Framework: A method to preserve privacy and perception. *Journal of Marketing Research*, 0022243720920256.

Zwebner, Y., Sellier, A. L., Rosenfeld, N., Goldenberg, J., & Mayo, R. (2017). We look like our names: The manifestation of name stereotypes in facial appearance. *Journal of Personality and Social Psychology*, 112(4), 527.