Audio Mining: The Role of Vocal Tone in Persuasion

XIN (SHANE) WANG
SHIJIE LU
XI LI
MANSUR KHAMITOV
NEIL BENDLE

Persuasion success is often related to hard-to-measure characteristics, such as the way the persuader speaks. To examine how vocal tones impact persuasion in an online appeal, this research measures persuaders' vocal tones in Kickstarter video pitches using novel audio mining technology. Connecting vocal tone dimensions with real-world funding outcomes offers insight into the impact of vocal tones on receivers' actions. The core hypothesis of this paper is that a successful persuasion attempt is associated with vocal tones denoting (1) focus, (2) low stress, and (3) stable emotions. These three vocal tone dimensions—which are in line with the stereotype content model—matter because they allow receivers to make inferences about a persuader's competence. The hypotheses are tested with a large-scale empirical study using Kickstarter data, which is then replicated in a different category. In addition, two controlled experiments provide evidence that perceptions of competence mediate the impact of the three vocal tones on persuasion attempt success. The results identify key indicators of persuasion attempt success and suggest a greater role for audio mining in academic consumer research.

Keywords: audio mining, machine learning, voice analytics, entrepreneurial pitches, persuasion, crowdfunding, artificial intelligence

Xin (Shane) Wang (xwang@ivey.uwo.ca) is an associate professor of marketing, Kraft Professor in Marketing at Ivey Business School, Western University, London, ON, Canada N6G 0N1. Shijie Lu is an assistant professor of marketing (slu@bauer.uh.edu) at Bauer College of Business, University of Houston, Houston, TX 77204. Xi Li (xili44@cityu.edu.hk) is an assistant professor of marketing at City University of Hong Kong, Kowloon Tong, Hong Kong. Mansur Khamitov is an assistant professor of marketing (mkhamito@iu.edu) at Kelley School of Business, Indiana University, Bloomington, IN 47405. Neil Bendle is an associate professor of marketing (neil.bendle@uga.edu) at Terry College of Business, University of Georgia, GA 30605. Please address correspondence to Xin (Shane) Wang. All authors contributed equally and are listed in reverse alphabetical order. The authors appreciate Nemesysco for proving the free access to voice analysis software QA5. The work was supported by the Social Sciences and Humanities Research Council (SSHRC) 435-2019-0597 and Indiana's RATS grant (1036260/MKHAM). Supplementary materials are included in the web appendix accompanying the online version of this article.

Editor: J. Jeffrey Inman

Associate Editor: Andrew T. Stephen

Advance Access publication February 23, 2021

A major stream of the consumer behavior literature involves understanding persuasion attempts. The profusion of online interactions provides avenues to study this important consumer-relevant phenomenon but also changes the process. For example, people who make persuasion attempts online (whom we call "persuaders") cannot be assessed through handshakes or eye contact. Yet, technology still gives the targets of persuasion (i.e., "receivers") cues that enable intuitive assessment. In particular, online videos allow receivers to hear the persuader's vocal tones. These provide cues that receivers can use to determine their response to the persuasion attempt.

In light of our interest in the ability of vocal tone measures to predict persuasion, we turn to the lens of the stereotype content model (SCM) (Fiske et al. 2002). We suggest that receivers use cues to determine whether the persuaders are likely to deliver what they promise. Specifically, we test whether persuaders' vocal tones, measured by a novel voice analysis software system, affect receivers' decisions

 • Vol. 48 • 2021
DOI: 10.1093/jcr/ucab012

to fund a request because vocal tones are thought to give insight into a persuader's competence. Our core idea is that speakers who seem more focused will be perceived as more competent, and thus more persuasive. In two studies, we examine videos that were posted online with the aim of securing relatively small amounts of funding from numerous nonexpert investors. In both of these studies, vocal tones showing signs of focus align with funding success. We follow our secondary data analysis with controlled experiments that allow us to demonstrate that the persuader's perceived competence mediates the receiver's response to the vocal tones.

PERSUASION AND VOCAL CHARACTERISTICS

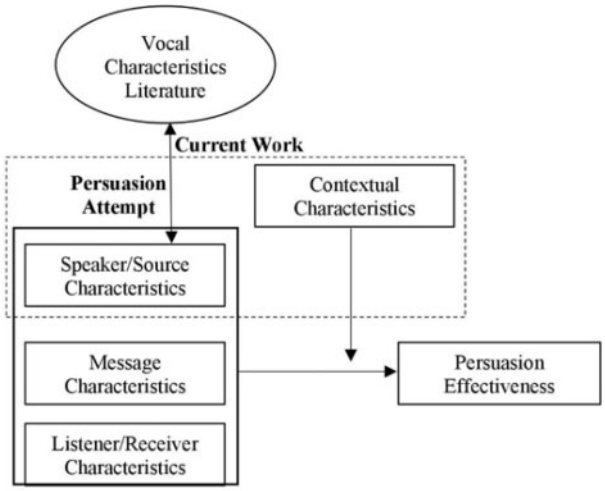
Vocal characteristics,[1] such as speech rate, pitch, volume, and tone, can influence perceptions of speaker traits and character. For example, lower-pitched voices are judged as signifying that the speaker is more potent/strong, competent, truthful, empathic, and trustworthy. Relatedly, faster speakers are typically judged to be more fluent, competent, socially attractive, truthful, and persuasive (Apple, Streeter, and Krauss 1979; Chattopadhyay et al. 2003; Cheng et al. 2016; Klofstad, Anderson, and Nowicki 2015; Klofstad, Anderson, and Peters 2012; Oleszkiewicz et al. 2017; Street, Brady, and Putman 1983; Tigue et al. 2012; Wiener and Chartrand 2014). Our findings connect to, and expand, this literature (Van Zant and Berger 2020), reinforcing that speaker characteristics matter—even when recorded and displayed online. We document how such characteristics can be inferred from vocal tones in online persuasion attempts and, critically, the effect these tones exert on persuasion in the marketplace in the form of real-world consequential funding outcomes. A central contribution of our work is to illustrate that online persuasion (Dillard and Shen 2013) can be mined at scale and captured with fine-grained detail using an automatic audio mining method. In doing so, we provide a pioneering account of how computers (as opposed to labor-intensive and/or semi-automated approaches to code and interpret voices) can predict an online persuasion attempt's effectiveness based on extracted vocal tone characteristics. Figure 1 and Table 1 present our research focuses in comparison with prior studies.

ONLINE PERSUASION AND AUDIO MINING

We observe persuasion on Kickstarter, a major online crowdfunding platform (Bayus 2013; Fan, Gao, and Steinhart 2020; Lin and Viswanathan 2016). Kickstarter straddles the commercial and noncommercial worlds and primarily caters

FIGURE 1

HOW THIS RESEARCH FITS WITH THE PERSUASION LITERATURE



to entrepreneurs/artists seeking support for their ideas. No equity stakes are given (Cholakova and Clarysse 2015) nor detailed business plans revealed, yet rewards are promised to those offering funding (e.g., a prerelease DVD). Kickstarter has been used in prior research (Mollick and Nanda 2016) to show that early support (Colombo, Franzoni, and Rossi Lamastra 2015), social connections (Mollick 2014), and geographical distance (Agrawal, Catalini, and Goldfarb 2014) impact funding. Importantly, the platform is not based on donations: the entrepreneurs/artists must persuade funders that they will deliver on the proposed project. Hence, intuition likely plays a vital role in this process, especially in making quick decisions. Receivers cannot always explain why they intuitively find some people more credible. However, one factor that facilitates intuitive judgments is hearing a persuader's vocal tones. Accordingly, we investigate exactly how these tones impact persuasion success.

Voice analysis has long proved challenging, but new technology can now quantify human voices relatively accurately. We collaborated with Nemesysco Ltd, an Israeli high-tech firm whose QA5 system (Nemesysco 2015) has been commercially applied in call centers and sold as "the most sophisticated, flexible, cutting edge voice analysis technology available today" (Nemesysco 2016). It not only allows analysis of many more audio files than could be achieved using human coders but also creates greater standardization (see web appendix A). Signal-processing algorithms extract and combine attributes from voices to identify, among other things, evidence of cognitive processes and of specific emotional reactions.

In addition to commercial uses in fraud detection, call center monitoring, and recruitment, audio mining technology has

1 We thank an anonymous reviewer and the Associate Editor for this suggestion.

TABLE 1

THIS RESEARCH VERSUS TRADITIONAL CONSUMER PERSUASION LITERATURE

Predominant focus	Most prior research	This research
Persuasion medium	Traditional, offline, and face-to-face	Online/digital persuasion
Persuasion effectiveness indicators	Perceptions, judgments, behavioral intentions, attitudes, small-scale behavior	Real-world consequential and tangible financial outcomes
Method to capture persuasion attempt/effectiveness	Manual, labor-intensive, and/or semi-automated	Automatically extracted, computationally mined, larger scale allied with experimental tests
Persuasion attempt factors	Message, listener/receiver, or speaker/source characteristics (e.g., source credibility, trustworthiness, type, attractiveness, similarity, topic expertise, and knowledge)	Speaker characteristics (e.g., focus, extreme emotion, stress); context; mediation through competence perceptions between characteristics and outcomes

featured in academic research. For instance, to predict the probability of a firm having to restate its accounts, one study searched 615 conference calls for indications of Chief Executive Officer deception (Hobson, Mayew, and Venkatachalam 2012). Similarly, 49 managers' emotional states on earnings calls conveyed information about firm financial prospects (Mayew and Venkatachalam 2012). But this past research used modest sample sizes and uncovered signals from trained communicators. In contrast, we examine a much larger sample of persuaders whose speaking voices largely lack professional training.

Audio mining technology analyzes numerous dimensions. Given our interest in the ability of QA5 measures to predict persuasion, we ground our research in the SCM (Fiske et al. 2002; Fiske, Cuddy, and Glick 2007). In this model, competence is one of two fundamental dimensions underlying impression formation, and hence, interpersonal persuasion. Defined as the extent to which an entity is perceived as capable (Fiske et al. 2002), competence relates to efficacy, skill, creativity, confidence, and intelligence (Cuddy, Fiske, and Glick 2008). Researchers have applied the SCM to branding (Fiske, Malone, and Kervyn 2012; Kervyn, Fiske, and Malone 2012), service interactions (Li, Chan, and Kim 2019), visual marketing (Wang et al. 2017), nonprofits (Aaker, Vohs, and Mogilner 2010), company size ratings (Yang and Aggarwal 2019), and communications (Dubois, Rucker, and Galinsky 2016).

We consider three vocal tones available in QA5 that are theoretically connected to perceived competence (or lack thereof). The SCM literature shows that higher competence is linked with being determined, being diligent, and focusing on a task (Cuddy et al. 2008, 72, 125). The QA5 vocal tone measure that best reflects the persuader exhibiting focus is labeled "concentration." Conversely, persuaders are judged as not confident in their own competence when they exhibit a lack of emotional regulation or evidence of stress in their voice (Wang et al. 2017). We can thus use the QA5 measures "extreme emotions" and "stress" to capture these two characteristics that suggest lack of competence.

Perceived competence impacts whether receivers predict persuaders will achieve their goals (Fiske et al. 2002, 2007; Kervyn et al. 2012). Indeed, why fund a project if you do not believe the persuader is competent to deliver it? Therefore, we expect that the perceived competence of persuaders mediates the effects of persuaders' vocal tones on funding success.

HYPOTHESES

Can the persuader deliver on the promises made? We suggest that receivers infer the answer to this question according to several dimensions of the persuaders' vocal tones. Our starting point is that we are interested in the ability of audio mining measures to predict persuasion. As noted, appearing focused and determined (i.e., "concentration" in QA5) may be seen as a proxy for persuader competence, a known driver of persuasion. Further indirect support for our prediction stems from research linking task engagement/involvement (i.e., focus) with a belief in competent and successful task execution (Csikszentmihalyi 1977; Elliot and Harackiewicz 1994; Nurttila, Ketonen, and Lonka 2015). If so, those conveying focus, after controlling for other relevant factors, will be more likely to be successful in their funding requests, given that inferred competence has been shown to predict dependent measures of campaign success, funding outcomes, vendor preference, and willingness to buy (Aaker et al. 2010; Guo et al. 2017; Zhou and Ye 2019). Thus, our first hypothesis is as follows:

> **H1:** Funding requests in video pitches are more likely to be successful when the voices of those requesting funds show greater signs of focus.

Successful results require skill as well as attention. Skill is difficult to judge but receivers can ask themselves: Do persuaders believe in their own ability and convey this confidence through their vocal tones (Kimble and Seidel 1991; Scherer, London, and Wolf 1973)? If not, receivers might observe telltale signs of stress in a persuader's voice

(Giddens et al. 2013; Hollien 1980). Although this is far from perfectly diagnostic—some competent people show signs of stress when speaking, whereas some lacking skill do not—stress in a persuader's voice may be perceived as revealing hidden information (Streeter et al. 1977), lowering perceived competence, and undermining the persuasion attempt (Apple et al. 1979). Indeed, extant research alludes to this possibility by underscoring that stressed entrepreneurs may be viewed as struggling to competently ensure project success (Grant and Ferris 2012; Wincent, Örtqvist, and Drnovsek 2008). Hence:

> **H2:** Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of stress.

A lack of realism in the objective may be conveyed by extreme levels of emotion in vocal tones (Carlo et al. 2012; Rafaeli and Sutton 1989; Shields 2005), which can imply that the persuader has embraced the project without adequately considering whether it is deliverable (Chen, Yao, and Kotha 2009). Receivers may therefore infer that excessively emotional persuaders are not competent to deliver a quality project. This link between extreme emotionality and lack of competence is hinted at in the work of Harker and Keltner (2001), who document that emotionality-related items like overreacting to minor frustrations or exhibiting the opposite of a calm and relaxed manner correlate negatively with ratings of competence. This leads to our third hypothesis:

> **H3:** Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of extreme emotion.

Beyond the isolated effects of vocal tones, there is a reason to expect an interactive effect between different vocal tones and videos featuring these voices on receivers. On the Kickstarter platform, funding requests are contained within videos (Koch and Siering 2019), some of which are more eye-catching than others. Stimulating videos tend to attract greater funding, which we attribute to gaining more attention (Li, Shi, and Wang 2019). We suggest that receivers are more influenced by the persuader's vocal tone when they are paying greater attention, as they do to relatively stimulating videos (Koch and Cheng 2016). In doing so, we build on prior research by Jiang and Benbasat (2007) arguing that stimulating video formats grab more attention of viewers because of the interactive interplay between dynamic scene changes and sound effects. As such, we do not expect vocal tone aspects to somehow clash or interfere with the video stimulation aspect. On the contrary, drawing on prior crowdfunding research on the interplay between video characteristics and spoken language, we expect the two to magnify each other (Cudmore and Slattery 2019; Korzynski, Haenlein, and Rautiainen 2021). Formally, we hypothesize that visual stimulation accentuates the impact of the vocal tone variables along the lines of a cross-modal influence on the direction of attentional processes (Krishna 2012; Krishna and Schwarz 2014). That is, to the extent that a video is stimulating, any effects of vocal tones—whether positive (i.e., signs of focus) or negative (i.e., signs of stress or extreme emotion)—should be exacerbated.

> **H4:** (a) Funding requests in video pitches are more likely to be successful when the voices of those requesting funds show greater signs of focus and the video is relatively stimulating.
>
> (b) Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of stress and the video is relatively stimulating.
>
> (c) Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of extreme emotion and the video is relatively stimulating.

Not all video features are expected to accentuate the impact of our focal vocal tones. Within videos, another important yet distinct form of contextual visual cue comes from the brightness of the picture (Zhang et al. 2019). Considered a simple and useful metric of visual information, brightness of a video is the average of the brightness/illumination of all its pixels (Li et al. 2019). Our logic, based on prior findings, is that excessive brightness can be off-putting, visually discomforting, and distracting (Ampenberger, Staggl, and Pohl 2017; Aylott and Mitchell 1998; Baker, Holland, and Kaufman-Scarborough 2007; Zhang et al. 2017). This logic is consistent with preliminary findings by Li, Shi, and Wang (2019), who recently documented a negative impact of high brightness levels on crowdfunding project success. If brightness interferes with and diverts attention (Custers et al. 2010; Proulx and Egeth 2007) from the funding request, we would expect this to attenuate the impact of the vocal tone. That is, we expect our core effects of vocal tones to manifest up to a certain, relatively acceptable level of brightness, after which any such effects should clash with and be diminished by excessive video brightness. Such a pattern of effects would also be consistent with, and directly build on, prior consumer behavior literature on extreme illumination in the context of retail store atmospherics and its overriding, attenuating impact (Baker, Levy, and Dhruv 1992; Mohan, Sivakumaran, and Sharma 2013; Summers and Hebert 2001).

> **H5:** (a) Funding requests in video pitches are more likely to be successful when the voices of those requesting funds show greater signs of focus and the video is less bright.
>
> (b) Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of stress and the video is less bright.

(c) Funding requests in video pitches are less likely to be successful when the voices of those requesting funds show greater signs of extreme emotion and the video is less bright.

We test our first five hypotheses on Kickstarter data and show associations between pitch characteristics and funding success. However, we would emphasize that these associations do not give us confidence of a causal explanation nor is perceived competence measured by the audio mining software. Therefore, we test an additional hypothesis in the laboratory—namely, that perceived competence (Cuddy et al. 2008; Fiske et al. 2002, 2007) will mediate the relationship between the vocal tone variables and outcome success. We ground our prediction in research suggesting that dimensions of vocal characteristics have a direct impact on person perception in the form of inferred competence and that such characteristics are key to competence judgments (Berry 1992; Berry et al. 1994; Brown, Strong, and Rencher 1975; Ray 1986; Rockwell 1996; Street et al. 1983; Street and Brady 1982). In turn, inferences of competence have been shown to drive outcome variables like campaign success, funding outcomes, and vendor preference (Aaker et al. 2010; Guo et al. 2017; Zhou and Ye 2019). Taken together:

H6: The extent to which greater signs of focus, stress, or extreme emotion in the vocal tones of those requesting funds predict the success of funding requests will be mediated by perceived persuader competence.

STUDY 1: KICKSTARTER FUNDING OF MUSIC PROJECTS

Study 1 uses audio mining technology to test hypotheses 1–5, determining the relationships between the characteristics of online video pitches and funding success.

Data

All data are from Kickstarter. Figure 2 illustrates a typical project on the Kickstarter platform, showing a video pitch right below the venture name.

We scraped all completed music projects, successful and unsuccessful, from three major markets (New York, Los Angeles, and Texas). Our raw data include 8327 projects from April 2009, Kickstarter's introduction, to December 2015, when we ended data collection. Unfortunately, 18.9% of projects no longer had valid video content; the video might have been withdrawn after project completion or faced server problems. This left a sample of 6755 projects. Table 2 presents the summary statistics.

We first consider the music project category because (1) it is one of the largest Kickstarter categories (Kickstarter 2020); (2) music projects are relatively homogenous, creating fewer concerns about confounding factors; and (3) music projects typically have the accompanying videos necessary for our audio analysis. (We replicate our findings within a different category in study 2.)

FIGURE 2

SNAPSHOT OF A KICKSTARTER VENTURE

TABLE 2

SUMMARY STATISTICS

	Variable name	Mean	SD	Min	Max
General	Success	.568	.495	0	1
	Target ($000s)	14.006	265.592	.001	21,475
	Collected ($000s)	5.672	13.489	0	600.874
	Project duration (days)	35.754	14.165	1	92
	Menu length	9.745	5.624	0	69
	First time	.809	.394	0	1
	Price	125.240	256.702	0	10,000
	FB_top 25%[a]	.141	.348	0	1
	FB_25–50%[a]	.135	.342	0	1
	FB_50–75%[a]	.142	.349	0	1
	FB_75–100%[a]	.126	.332	0	1
	Solo	.689	.463	0	1
	Female (solo artist)	.234	.424	0	1
Text	Word count	499.242	371.695	17	6546
	Positive %	.028	.012	0	.111
	Negative %	.006	.006	0	.055
Audio test variables	Focus	6.560	2.602	0	30
	Stress	6.969	3.758	0	26
	Extreme emotion	2.930	1.670	0	16
Audio control variables	Content	.647	1.700	0	26
	Excitement	17.430	3.379	0	30
	Angry	.509	1.225	0	17.902
	Imagination activity	1.802	1.694	0	13.506
Video	Video duration (s)	201.677	120.974	8	2089
	Visual stimulation	.520	.245	0	1.510
	Brightness	101.475	32.256	10.659	252.565
Genre	Classical	.055	.228	0	1
	Country	.072	.258	0	1
	Hip-hop	.064	.245	0	1
	Electronic	.033	.179	0	1
	Jazz	.055	.229	0	1
	Pop	.089	.285	0	1
	Rock	.123	.329	0	1
	World	.042	.200	0	1
	Other	.466	.499	0	1
Observations			6755		

[a]Compared to those without reported Facebook accounts.

Dependent Variable Our dependent variable is whether the venture was funded. Kickstarter projects set a target and funds are only released if this target is reached, giving a clear success metric. Those that do not reach their target receive nothing. Thus, *Success* is a binary variable taking the value one when the amount of funds pledged—the variable *Collected*—is greater than the target amount, and zero otherwise. As table 2 shows, 56.8% of ventures in our data were successfully funded, receiving an average of $5672 each and a maximum of $600,000.

Independent Variables. Our key independent variables relate to the persuader's vocal characteristics, as measured by the QA5 software. All variables generated are automatically standardized with a value between 0 and 30. We analyze three focal variables:

- *Focus* (H1): Concentration in QA5, indicating how task-focused the speaker sounds.
- *Stress* (H2): Indicates how nervous the speaker sounds.
- *Extreme emotion* (H3): Indicates how extreme the overall emotional activity is.

As these measures are detected by the QA5 software, they inevitably have a somewhat black box nature. To better understand the measures, we drilled into the method to determine the precise signals underpinning the QA5 measures. We detail this analysis in web appendix A.

Other vocal tone variables provided by QA5 are:

- *Content:* Indicates how pleased or happy the speaker sounds.
- *Excitement:* Indicates how positively or negatively excited the speaker sounds.
- *Angry:* Indicates how angry the speaker sounds.
- *Imagination activity*: Indicates the extent to which the speaker sounds like they are imagining rather than recalling information.

Control Variables We control for factors commonly considered by investors and prior research. As noted, a project is funded only if the amount pledged, the variable *Collected*, is greater than, or equal to, the target. Given its

importance to success, we control for the *Target* set by the artist at project initiation; in this particular study, we also refer to the persuader as the "artist." As *Target*'s distribution is widely dispersed, we use its natural log to mitigate the impact of potential outliers. Success is assessed at the end of the prespecified number of days, *Project duration*, after which the venture is closed and can receive no funds. *Menu length* is the number of reward options offered by the artist, which include gratitude, a prerelease download, a video, or album credit; Hu, Li, and Shi (2015) show that this measure relates to funding success. *Price* is the median amount that the funder has to commit to receive a reward. We also create dummy variables for the artist's number of Facebook friends (Mollick 2014); for example, FB top 25% equals one if an artist is in the top quartile with respect to number of friends.

The binary variable *First time* indicates whether the artist is new to Kickstarter; 81% had not submitted previously. We observe music genre (i.e., *Genre* can be classical, country, hip-hop, electronic, jazz, pop, rock, world music, or other). We infer whether the artist is a solo artist or a band/group of artists from IDs using a machine learning algorithm. Specifically, the Python package "SexMachine" recognizes the artist's gender using a first name dictionary (e.g., Julia and Serena are considered female, Jay and Brian male). Unrecognized names and groups identified as "unknown" are mostly band names (e.g., "Chocolatestar Music"). Therefore, we treated female and male names as solo artists and unrecognized ones as band names. Solo artists represent 69% of projects; female solo artists provide 23% of projects and male artists 46%.

To control for the accompanying text, we total its words, *Word count*. We also conduct a sentiment analysis using the positive (2006) and negative (4783) word lists complied by Hu and Liu (2004) and Liu, Hu, and Cheng (2005). *Positive%* and *Negative%* are the text's valence.

We further control for video characteristics using measures related to *Video duration*, *Visual stimulation*, and *Brightness*. *Visual stimulation* is captured by the average difference in images between frames; higher differences indicate more stimulation. Li et al. (2019) develop a method to automatically measure the visual stimulation of videos, showing that stimulation has a significant positive impact on funding. Following their method, we divide a video into 10 equal-distanced clips and then compute the visual difference between the 10 frames using the frame in the middle of each clip. To measure brightness, we convert a frame into a grayscale image and calculate the root mean square (RMS) of the grayscale of all pixels. We then average the RMS brightness over video frames to calculate the entire video's brightness.

Method

To examine the relationship between vocal tones and the success of a funding request, we estimate the following Probit model:

$$\Pr(\text{Success}_i) = \Phi(\beta_0 + \beta_1 X_i^{\text{General}} + \beta_2 X_i^{\text{Text}} + \beta_3 X_i^{\text{Voice}} + \beta_4 X_i^{\text{Image}} + \text{Genre}_i + \text{Year}_i), \quad (1)$$

where Success_i is a binary indicator equaling one if the project i is fully funded, Φ is the cumulative density function of a standard normal distribution, X_i^{General} includes project characteristics: log of target amount, project duration, menu length, price, artist's demographics (solo/gender), Kickstarter experience, and Facebook activity, X_i^{Text} captures the accompanying text: total words and sentiment, X_i^{Voice} includes the audio variables for our hypotheses, and X_i^{Image} represents video length, stimulation, and brightness. Last, we include both genre and year dummies (i.e., when posted) to control for potential heterogeneity across music genres and time.

Results

Equation (1)'s estimation results are reported in the first column of table 3, Model 1.

The estimated coefficients of characteristics all have expected signs. First-time projects are more likely to be funded than repeat projects, as are smaller targets compared to more ambitious goals. Bands are more likely to be funded than solo artists, with female solo artists more likely to be funded than males. Artists with Facebook friends above the lowest 25% are more successful, whereas those in the lowest 25% are less successful than artists without a reported account. More text in the accompanying project description is positively associated with success, as is a more neutral sentiment in the text (compared to a more positive or negative sentiment). In a possible parallel to hypothesis 3 (regarding the presence of extreme emotion in vocal tone), we might expect competence to be conveyed by a neutral written sentiment. Finally, projects with shorter, less bright, and more stimulating videos gain greater success. The first column of Model 1 shows that artist focus level, as measured by the audio mining software, is positively associated with funding success and therefore supports hypothesis 1. Our results indicate that a project is indeed more likely to be funded when an artist has a less-stressed vocal tone, supporting hypothesis 2. Our results also show a statistically significant and negative relationship between extreme emotion and funding, supporting hypothesis 3.

As shown in the second column of table 3, Model 2, the effects of focus and extreme emotion are positively moderated by visual stimulation and negatively moderated by brightness, suggesting that hypotheses 4a, 4c, 5a, and 5c are supported. The interaction effect of stress and visual stimulation is negative, as predicted by hypothesis 4b, but not statistically significant. Similarly, the interaction effect of stress and brightness is positive, as predicted by

TABLE 3

RESULTS FROM THE PROBIT MODEL OF FUNDING SUCCESS—MUSIC CATEGORY

	DV = Success	Model 1 (no interactions)	Model 2 (interactions)
General attributes	Intercept	**2.773 (.236)**	**2.897 (.243)**
	Log(Target)	**−.324 (.021)**	**−.328 (.021)**
	Project duration	**−.012 (.001)**	**−.012 (.001)**
	Menu length	**.068 (.004)**	**.068 (.005)**
	First time	**.237 (.044)**	**.238 (.044)**
	Price	−.00018 (.0001)	−.00017 (.0001)
	FB_top 25%	**.368 (.055)**	**.369 (.055)**
	FB_25–50%	**.124 (.053)**	**.133 (.054)**
	FB_50–75%	**.311 (.053)**	**.310 (.053)**
	FB_75–100%	**−.244 (.055)**	**−.243 (.055)**
	Solo	**−.148 (.040)**	**−.154 (.040)**
	Female	**.340 (.044)**	**.340 (.045)**
Text	Word count	**.0005 (7e−5)**	**.0005 (7e−5)**
	Positive %	**−3.252 (1.440)**	**−3.239 (1.442)**
	Negative %	**−23.704 (2.997)**	**−23.834 (2.999)**
Audio test variables	Focus (H1)	**.019 (.007)**	**.025 (.008)**
	Stress (H2)	**−.016 (.006)**	**−.018 (.006)**
	Extreme emotion (H3)	**−.052 (.016)**	**−.063 (.016)**
Audio control variables	Content	.015 (.014)	.023 (.015)
	Excitement	−.005 (.005)	−.005 (.005)
	Angry	.002 (.017)	.007 (.018)
	Imagination activity	.013 (.013)	.020 (.014)
Video	Video duration	**−.0005 (.0001)**	**−.0006 (.0001)**
	Visual stimulation	**.828 (.076)**	**.815 (.077)**
	Brightness	**−.002 (.0006)**	**−.003 (.0008)**
Interaction effects with visual stimulation	Focus × visual stimulation (H4a)		**.074 (.029)**
	Stress × visual stimulation (H4b)		−.011 (.025)
	Extreme emotion × visual stimulation (H4c)		**−.200 (.059)**
	Content × visual stimulation		**.156 (.060)**
	Excitement × visual stimulation		.015 (.023)
	Angry × visual stimulation		.054 (.071)
	Imagination × visual stimulation		**.160 (.054)**
Interaction effects with brightness	Focus × brightness (H5a)		**−.0004 (.0002)**
	Stress × brightness (H5b)		.0002 (.0002)
	Extreme emotion × brightness (H5c)		**.0008 (.0004)**
	Content × brightness		−.0006 (.0004)
	Excitement × brightness		.0002 (.0002)
	Angry × brightness		−.0002 (.0005)
	Imagination × brightness		−.0006 (.0004)
Additional controls	Genre fixed effects	Yes	Yes
	Year fixed effects	Yes	Yes
Summary	Observations	6755	6755
	Pseudo R^2	.197	.200
	Log-likelihood	−3711.2	−3694.1

Notes. Coefficient estimates significant at 95% level are in bold. All interaction terms are created using mean-centered variables. Robust standard errors are in parentheses.

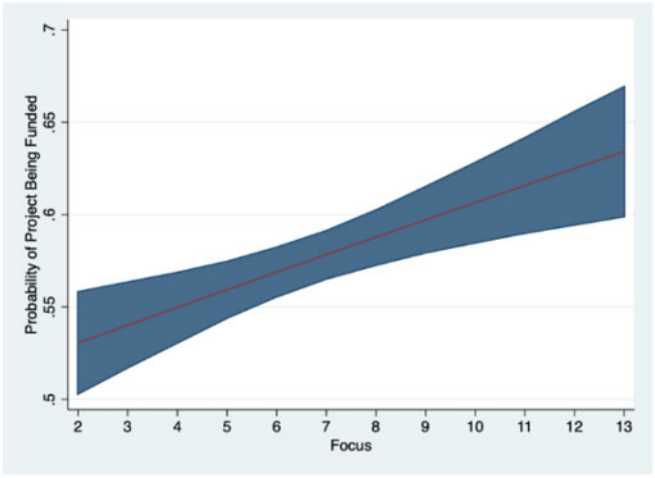
hypothesis 5b, but statistically nonsignificant. As a robustness check, we also tested hypotheses 1–5 using a continuous (i.e., log(Collected)) rather than binary (i.e., Success) dependent variable. Further, we reran the analyses for both Studies 1 and 2 using proportion of funding collected. We find substantively similar results and report the details in web appendix B.

To calibrate the effects of focus, stress, and extreme emotion on funding, we specified the range of each variable from its values at the 2.5 and 97.5 population percentiles. For example, the *X*-axis of *Focus* in figure 3*A* ranges from 2 to 13 because its 95% interval is from 2.30 to 12.58. Using coefficient estimates from Model 2, we depict the relationship between success and vocal tone, keeping other covariates at the mean level. All three effects are substantial. The least focused, most stressed, and most emotionally extreme persuaders (i.e., on the negative extreme of the *X*-axes in figure 3) would have had a greater chance of success by about 10, 8, and 14 points, respectively, if they were on the positive extreme.

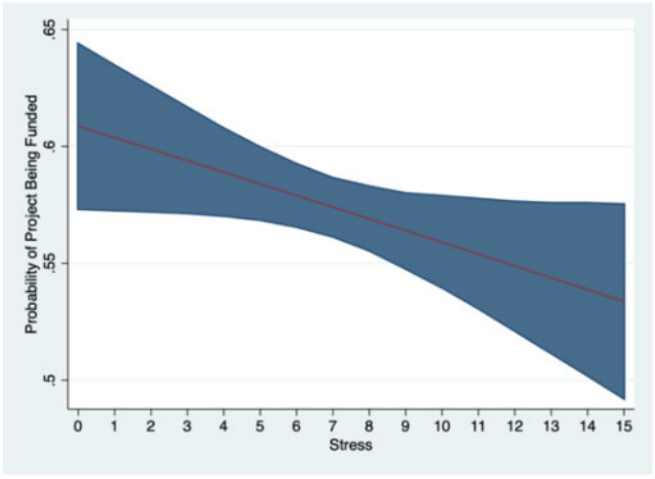
Although we had no a priori hypothesis here, the significant and positive coefficient of *Content* ×

FIGURE 3

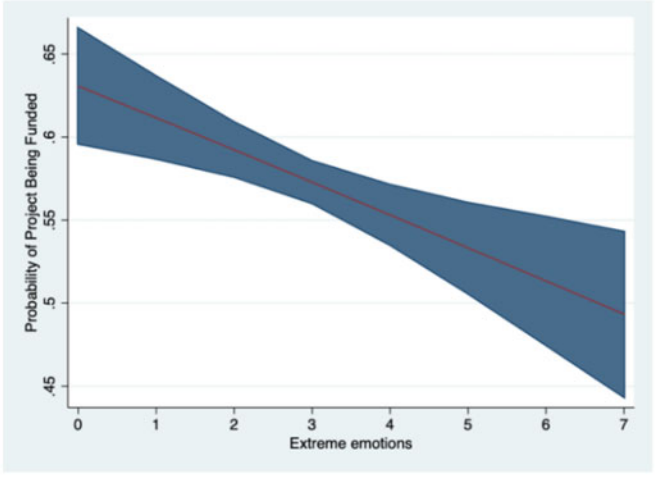
RELATIONSHIP BETWEEN PREDICTED FUNDING SUCCESS RATE AND VOCAL TONE.



(**a**) Focus



(**b**) Stress



(**c**) Extreme emotion

NOTES. DARK AREAS CORRESPOND TO 95% CONFIDENCE INTERVAL OF PREDICTIONS

Stimulation and *Imagination* × *Stimulation* may mean happier tones with a low degree of information recalling are more successful when the video is also more stimulating. For completeness, we investigated other potential moderators, including target amount, project duration, and video duration, but did not find any significant evidence of interactions.

Discussion of Study 1

Study 1 reveals a significant relationship between focus, lack of stress, and lack of excessive emotionality on funding success. The results presented in table 3 lend support for hypotheses 1, 2, 3, 4a, 4c, 5a, and 5c, but not hypotheses 4b and 5b. Yet music projects, though a large and interesting category, represent but one type of funding project on Kickstarter. Accordingly, we next conduct a study in a different category in an attempt to replicate the findings from study 1.

STUDY 2: REPLICATION IN A NONMUSIC SETTING

Data Description and Method

In study 2, we replicate most study 1 findings using a different Kickstarter category, technology. Like music, the technology category is one of the largest on Kickstarter, providing us with sufficient observations to test our hypotheses. We collected data on all technology projects in six U.S. states: California, Illinois, Massachusetts, New York, Texas, and Washington. The data set comprises all completed projects from April 2009 to March 2017. Following the sampling method in study 1, we focus on technology projects with a video pitch, leaving study 2 with 3966 observations (75.5% of 5252 technology projects). The variables used in study 2 are the same as those used in study 1 except we replace *Genre*, used for music projects, with *Type*, which refers to the type of the technology project (i.e., gadgets, hardware, do-it-yourself electronics, flight, 3D printing, apps, camera equipment, or other).

Results

Table 4 presents the estimation results from Equation (1) using technology projects. We find that, as in study 1, *Focus* is positively associated with funding success, while *Extreme emotion* is negatively associated with the same—providing further support for hypotheses 1 and 3. However, hypothesis 2 is not supported, as stress does not seem to be associated with funding success in the technology category. A possible explanation for this result is that sounding stressed may be less diagnostic for technology projects because they are relatively complex and technical (Meuter et al. 2003; Mick and Fournier 1998) and thus stressful for an average Kickstarter entrepreneur to convey effectively. For hypotheses 4 and 5, results of the model with interactions (Model 2) indicate that *Focus* interacts with both visual stimulation (positively) and brightness (negatively), whereas other interactions do not show an effect. (Web appendix B shows that these findings hold running the ordinary least squares (OLS) regression of Log(*Target*) on the covariates. We also provide robustness checks for Studies 1 and 2 using a different software in Web Appendix C.)

The effect of *Focus* is most generalizable (i.e., hypotheses 1, 4a, and 5a are supported in both studies). The effect of *Stress* is least generalizable (i.e., hypothesis 2 is supported only in study 1). *Extreme emotion's* effect lies in between in terms of generalizability (i.e., hypothesis 3 is supported in both studies, whereas hypotheses 4c and 5c gain support only in study 1).

Thus, using data sets from two Kickstarter categories, we find a consistent impact of information in the speaker vocal tones (especially *Focus*) on funding outcomes. In these ecologically valid studies, however, we do not observe receivers' perception of persuaders' competence, which prevents us from testing the mediation. We therefore turn to experimental setups to show that perceptions of competence mediate the relationship between vocal tone and funding success.

STUDY 3: EXPERIMENTAL VALIDATION AND MECHANISM TEST

The first objective of study 3 is to experimentally investigate and validate the findings of Studies 1 and 2. Consistent with hypotheses 1–3, we expect that a funding request is less likely to succeed when the voice requesting funds in a video pitch shows greater signs of stress or extreme emotion, and more likely to succeed when the voice shows greater signs of focus.

The second goal of study 3 is to shed light on the mechanism underlying the impact of vocal tone on the success of funding requests (H6). We predict that exposing receivers to pitches, where the voice of the persuader exhibits greater signs of extreme emotion or stress may reduce perceived persuader competence, whereas greater signs of focus may enhance it. Perceived persuader competence then results in greater success of a funding request. Figure 4 shows a visual representation of the conceptual framework of the current research.

The third goal of study 3 is to rule out alternative explanations for our findings. Beyond perceived persuader competence, other psychological mechanisms might also explain the core effect. For example, people tend to like individuals who appear determined and focused more than those who do not (Asch 1946; Fiske and Neuberg 1990), and liking, in turn, is associated with greater processing fluency (Hildebrand et al. 2017; Reber, Winkielman, and

TABLE 4

RESULTS FROM THE PROBIT MODEL OF FUNDING SUCCESS—TECHNOLOGY CATEGORY

	DV = Success	Model 1 (no interactions)	Model 2 (interactions)
General attributes	Intercept	**1.518 (.303)**	**1.559 (.308)**
	Log(Target)	**−.365 (.019)**	**−.367 (.019)**
	Project duration	.0005 (.002)	.0006 (.002)
	Menu length	**.071 (.007)**	**.071 (.007)**
	First time	**.366 (.054)**	**.367 (.054)**
	Price	**.0002 (.00004)**	**.0002 (.00005)**
	FB_top 25%	**.237 (.075)**	**.235 (.075)**
	FB_25–50%	.126 (.073)	.117 (.073)
	FB_50–75%	.101 (.077)	.096 (.078)
	FB_75–100%	**−.321 (.080)**	**−.326 (.080)**
	Solo	**−.511 (.051)**	**−.508 (.051)**
	Female	**.392 (.088)**	**.389 (.089)**
Text	Word count	**.0003 (.00003)**	**.0003 (.00003)**
	Positive %	**8.797 (2.231)**	**8.678 (2.233)**
	Negative %	**.474 (3.284)**	.119 (3.287)
Audio test variables	Focus (H1)	**.022 (.008)**	**.023 (.008)**
	Stress (H2)	.007 (.008)	.006 (.008)
	Extreme emotion (H3)	**−.055 (.023)**	**−.052 (.024)**
Audio control variables	Content	.041 (.023)	.019 (.026)
	Excitement	.005 (.007)	.003 (.007)
	Angry	−.003 (.029)	.010 (.032)
	Imagination activity	.029 (.015)	.027 (.015)
Video	Video duration	.0002 (.0002)	.0002 (.0002)
	Visual stimulation	**1.055 (.101)**	**1.046 (.103)**
	Brightness	−.001 (.0006)	−.0009 (.0006)
Interaction effects with visual stimulation	Focus × visual stimulation (H4a)		**.068 (.031)**
	Stress × visual stimulation (H4b)		−.007 (.029)
	Extreme emotion × visual stimulation (H4c)		.018 (.091)
	Content × visual stimulation		.027 (.102)
	Excitement × visual stimulation		−.030 (.027)
	Angry × visual stimulation		−.001 (.128)
	Imagination × visual stimulation		.065 (.056)
Interaction effects with brightness	Focus × brightness (H5a)		**−.0004 (.0002)**
	Stress × brightness (H5b)		−.00002 (.0002)
	Extreme emotion × brightness (H5c)		−.0007 (.0006)
	Content × brightness		−.001 (.0007)
	Excitement × brightness		.00004 (.0002)
	Angry × brightness		−.0001 (.0008)
	Imagination × brightness		−.0004 (.0004)
Additional controls	Type fixed effects	Yes	Yes
	Year fixed effects	Yes	Yes
Summary	Observations	3966	3966
	Pseudo R^2	.258	.261
	Log-likelihood	−1949.8	−1941.5

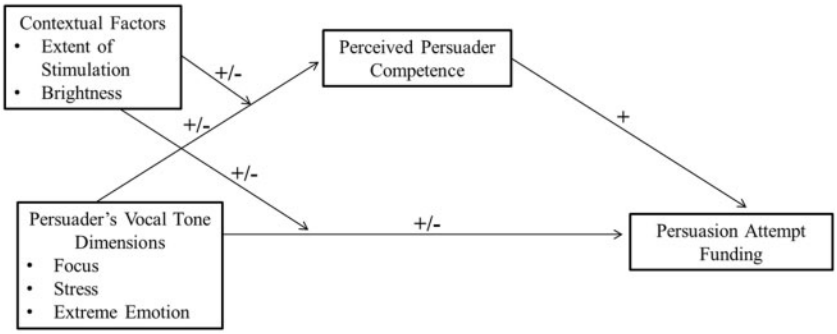
Notes. Coefficients significant at 95% level are in bold. All interaction terms are created using mean-centered variables. Robust standard errors are in parentheses.

Schwarz 1998; Winkielman and Cacioppo 2001). Given this link, it is plausible that receivers may more fluently process a pitch from a focused-sounding (vs. a control) speaker. Additionally, individuals who sound stressed or extremely emotional might leave receivers with a feeling that something does not add up, lowering processing fluency. Levels of processing fluency could then have a subsequent impact on how much receivers like (or dislike) a pitch, and by extension, the success of a funding request.

Another concern is trust (Swan et al. 1988; Swan, Bowers, and Richardson 1999). Receivers may not trust persuaders who appear distressed or extremely emotional (insecure) but rather, trust those who appear focused (secure). These levels of trust (Chaudhuri and Holbrook 2001; Khamitov, Wang, and Thomson 2019; Morgan and Hunt 1994) or distrust in a persuader may guide a peer-to-peer funding decision. In summary, study 3 examines two alternative mediating mechanisms: processing fluency and perceived persuader trust.

FIGURE 4

CONCEPTUAL FRAMEWORK OF CURRENT RESEARCH



Method

Participants, Design, and Procedure. We randomly assigned 413 Amazon Mechanical Turk (MTurk) volunteers (51% females, $M_{age} = 32.4$) to one of four conditions (vocal tone dimension: control vs. extreme emotion vs. stress vs. concentration) in a single-factor between-subjects design in exchange for a small amount of financial compensation. To increase the ecological validity of our stimuli, we conducted an extensive search to identify real-world pitches on crowdfunding platforms. Thus, the pitch was adopted from an actual crowdfunding campaign. The study was introduced as being about a new brand that was being pitched on Kickstarter. Participants were presented with an audio recording of a pitch for "COOLEST Cooler," a multifunction cooler brand. The pitch was recorded by an independent and qualified female research assistant blind to the experimental hypotheses who posed as the entrepreneur seeking funds.

Manipulating Vocal Tone. In manipulating vocal tone dimensions, we follow prior research on acoustics and phonetics (Aucouturier et al. 2016; Boidron et al. 2016; Rachman et al. 2018), which demonstrates that auditory perception of speech sounds can be predictably manipulated and altered. To avoid confounding effects, in all conditions, the pitch text and length was constant (i.e., "The COOLEST is a portable party disguised as a cooler, bringing blended drinks, music, and fun to any outdoor occasion…"; web appendix D details the full text). The only difference lay in how the pitch was manipulated to sound in terms of the vocal tone dimensions. In the control condition, participants listened to a pitch that was made to sound neutral and regular, using an everyday, natural tone. In the stress condition, we altered the pitch so the speaker sounded anxious, tense, distressed, and nervous. In the extreme emotion condition, we modified the pitch to sound overly excited, loud, and unnaturally energetic—as if the speaker was failing to control her emotions. Last, in the focus condition, we manipulated the pitch so that the speaker would sound focused, careful, and diligent.

After listening to the audio pitch, participants first responded by providing the dependent measures of "amount willing to fund the campaign" and brand evaluations. Importantly, they were likely to believe that their responses would have consequential effects.[2] They then answered questions measuring the vocal tone dimensions, which served as manipulation checks, as well as a perceived pitch length question (a pitch perceived as shorter vs. longer is indicative of speech rate/speed perceptions, i.e., how fast vs. slow the persuader's speech sounds). Next, respondents provided their ratings of the focal mediator: perceived persuader competence. They also provided alternative process variables (processing fluency and perceived persuader trust). The study ended with a brief demographic section, a suspicion probe, and thanks.

Measures. The amount each participant would be willing to fund was assessed on a 0-to-100 slider scale: "I am willing to support the COOLEST Cooler brand by contributing \$0–100" and "I am 0–100% more willing to fund the COOLEST Cooler brand compared to other cooler brands." These were combined to form a willingness to fund scale ($\alpha = .80$). Brand evaluations were measured on a seven-point semantic differential scale (i.e., unpleasant–pleasant; bad–good; negative–positive; unfavorable–favorable; dislike–like; useless–useful; not beneficial–

2 Participants were told that they would be asked to listen and evaluate a new brand being pitched on Kickstarter. After listening to the recording, participants were asked how much they were willing to fund/support the brand based on the pitch they heard. Participants arguably thought, at the moment when they were responding, that the researcher was capturing their actual willingness to fund. Upon completion of the study, we made sure to debrief respondents that no money would be solicited/requested from them by either the research team or Kickstarter.

beneficial; worthless–valuable; disagreeable–agreeable; $\alpha = .96$; Batra and Stayman 1990).

Manipulation checks pertaining to vocal tone dimensions were assessed by having respondents indicate the following items: the extent to which the voice of the persuader sounded neutral, regular, everyday sounding, ordinary (control check; $\alpha = .78$), extremely emotional, overly excited, unnaturally loud, emotionally unstable (signs of extreme emotion check; $\alpha = .74$), stressed, anxious, tense, nervous (signs of stress check; $\alpha = .92$), focused, concentrated, determined, attentive, and mindful (signs of focus check; $\alpha = .84$). Answers were collected on scales ranging from 1 (*not at all*) to 7 (*very*). A perceived pitch length check was administered on a seven-point semantic differential scale (i.e., short–long).

Perceived persuader competence was assessed with six items: the speaker (1) has the ability to convey her intentions to ordinary people, (2) is skilled and effective at achieving her goals, (3) seems to be competent, (4) seems to be capable, (5) seems to be confident, and (6) seems to be self-assured. Responses were collected on scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). These six items were combined in a perceived persuader competence scale ($\alpha = .95$), (Fiske et al. 2002; Kervyn et al. 2012).

Last, we measured processing fluency and perceived trust in the person making the pitch. Processing fluency was assessed with three items: (1) How easy was this information to comprehend? (2) How difficult was this information to understand (reverse coded)? and (3) How easy was this information to process ($\alpha = .92$; Lee and Aaker 2004)? Perceived persuader trust was captured with four items: (1) I trust this speaker, (2) I rely on this speaker, (3) this is an honest speaker, and (4) the speaker is safe ($\alpha = .88$; Chaudhuri and Holbrook 2001).

Results

Manipulation Checks The results of a one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. concentration) multivariate analysis of variance (MANOVA) with the four respective indices of vocal tone dimensions as dependent variables revealed that our manipulations were successful: control ($F(3, 409) = 9.460$, $p = .001$, $\eta^2 = .065$), extreme emotion ($F(3, 409) = 4.962$, $p = .002$, $\eta^2 = .035$), stress ($F(3, 409) = 59.202$, $p = .001$, $\eta^2 = .303$), and focus ($F(3, 409) = 16.669$, $p = .001$, $\eta^2 = .109$). Participants in the control condition reported higher scores on the control check compared to respondents in the extreme emotion ($M_{control} = 4.58$ vs. $M_{extreme\ emotion} = 3.73$, $p = .001$), stress ($M_{control} = 4.58$ vs. $M_{stress} = 3.70$, $p = .001$) or focus ($M_{control} = 4.58$ vs. $M_{focus} = 4.11$, $p = .013$) conditions. Similarly, participants in the extreme emotion condition reported elevated signs of extreme emotion compared to respondents in the control ($M_{extreme\ emotion} = 2.98$ vs. $M_{control} = 2.26$, $p = .001$), stress ($M_{extreme\ emotion} = 2.98$ vs. $M_{stress} = 2.57$, $p = .031$), and focus ($M_{extreme\ emotion} = 2.98$ vs. $M_{focus} = 2.56$, $p = .026$) conditions. Next, participants in the stress condition rated higher signs of stress compared to respondents in the control ($M_{stress} = 4.95$ vs. $M_{control} = 2.52$, $p = .001$), extreme emotion ($M_{stress} = 4.95$ vs. $M_{extreme\ emotion} = 2.60$, $p = .001$), and focus ($M_{stress} = 4.95$ vs. $M_{focus} = 2.65$, $p = .001$) conditions. Finally, participants in the focus condition generated stronger signs of focus ratings compared to respondents in the control ($M_{focus} = 5.38$ vs. $M_{control} = 4.84$, $p = .001$), extreme emotion ($M_{focus} = 5.38$ vs. $M_{extreme\ emotion} = 4.84$,[3] $p = .002$), and stress ($M_{focus} = 5.38$ vs. $M_{stress} = 4.19$, $p = .001$) conditions. Importantly, there were no significant differences in terms of perceived pitch length ($F(3, 409) = 1.256$, $p = .289$, $\eta^2 = .009$), suggesting that speech rate/speed perceptions were unaltered by our manipulations.

We conducted a post-test using the same population from the main study to confirm that our manipulations did not influence alternative factors (e.g., perceptions of attractiveness or age).[4] We randomly assigned 272 respondents (55% females, $M_{age} = 32.7$) to one of four conditions using the manipulations from the study. We adapted our measures from Addington (1968) and Collins and Missing (2003) to capture perceived speaker attractiveness (ugly–good-looking; unattractive–attractive; $\alpha = .92$) and speaker's perceived age (old–young; mature–youthful; $\alpha = .75$). The vocal tone dimension manipulation had no effect on either perceived speaker attractiveness ($M_{control} = 4.76$ vs. $M_{extreme\ emotion} = 4.76$ vs. $M_{stress} = 4.67$ vs. $M_{focus} = 5.00$; $F(3, 268) = .853$, $p = .466$, $\eta^2 = .009$; p's $> .14$) or age ($M_{control} = 5.70$ vs. $M_{extreme\ emotion} = 5.84$ vs. $M_{stress} = 5.75$ vs. $M_{focus} = 5.99$; $F(3, 268) = .701$, $p = .552$, $\eta^2 = .008$; p's $> .17$). This result suggests that our manipulation of vocal tone dimensions did not (significantly) impact perceptions of speaker attractiveness or age.

Amount Willing to Fund and Brand Evaluations A one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. focus) MANOVA yielded a significant multivariate main effect of vocal tone dimensions on dependent measures of amount willing to fund and brand evaluations ($F(6, 816) = 10.74$, Wilk's $\lambda = .862$, $p = .001$, $\eta^2 = .072$).

Follow-up univariate ANOVAs examined the nature of the main effect on each dependent variable. As hypothesized, a one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. focus) ANOVA with amount willing to fund as our dependent variable revealed a main effect of vocal tone dimensions ($F(3, 409) = 10.630$, $p <$

3 The actual means and SDs on the focus check were $M_{control} = 4.837$, SD = 1.31 versus $M_{extreme\ emotion} = 4.840$, SD = 1.07.

4 We would like to thank an anonymous reviewer for making this suggestion.

TABLE 5

CELL MEANS IN STUDY 3

	Condition			
	Control ($n = 103$)	Focus ($n = 107$)	Stress ($n = 103$)	Extreme emotion ($n = 100$)
Amount willing to fund	32.14 (30.03)	41.93 (30.66)*	23.68 (27.29)*	22.80 (23.08)*
Brand evaluations	5.38 (1.37)	5.73 (1.05)**	4.50 (1.54)**	4.85 (1.58)**
Perceived competence	5.04 (1.44)	5.56 (1.11)*	3.29 (1.55)*	4.69 (1.35)**
Processing fluency	5.45 (1.54)	5.43 (1.46)	5.11 (1.40)**	5.35 (1.40)
Perceived persuader trust	4.64 (1.44)	4.61 (1.57)	4.43 (1.40)	4.53 (1.13)

Notes. Standard deviations in parentheses;
*a significant difference between a control and a vocal tone dimension at $p < .05$;
**a marginally significant difference at $p < .10$.

.001, $\eta^2 = .072$). As table 5 shows, sounding focused elicited 30.5% higher "amount willing to fund" than its control condition counterpart ($M_{focus} = 41.93$ vs. $M_{control} = 32.14$, $p = .012$), further supporting Hypothesis 1. In addition, vocal tones that displayed signs of stress ($M_{stress} = 23.68$ vs. $M_{control} = 32.14$, $p = .031$) or extreme emotion ($M_{extreme\ emotion} = 22.80$ vs. $M_{control} = 32.14$, $p = .018$) lowered the amount willing to fund compared to the control condition (table 5); this represents a decrease of 26.3% and 29.1% in the amount willing to fund due to signs of stress and extreme emotion, respectively, in vocal tones, further supporting hypotheses 2 and 3.

We visualized the difference across conditions. Figure 5*A* shows a significantly increased willingness to fund persuaders who sound focused and a significantly decreased willingness to fund those who sound stressed or extremely emotional. Parallel results emerged with respect to brand evaluations, as figure 5*B* depicts. A one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. focus) follow-up univariate ANOVA with brand evaluations as the dependent variable yielded a main effect of vocal tone dimensions ($F(3, 409) = 16.006$, $p < .001$, $\eta^2 = .105$). Compared to the control condition, the focused vocal tone generated marginally enhanced brand evaluations ($M_{focus} = 5.73$ vs. $M_{control} = 5.38$, $p = .072$). Both the stressed vocal tone ($M_{stress} = 4.50$ vs. $M_{control} = 5.38$, $p = .001$) and the extremely emotional vocal tone ($M_{extreme\ emotion} = 4.85$ vs. $M_{control} = 5.38$, $p = .007$) showed significantly decreased brand evaluations.

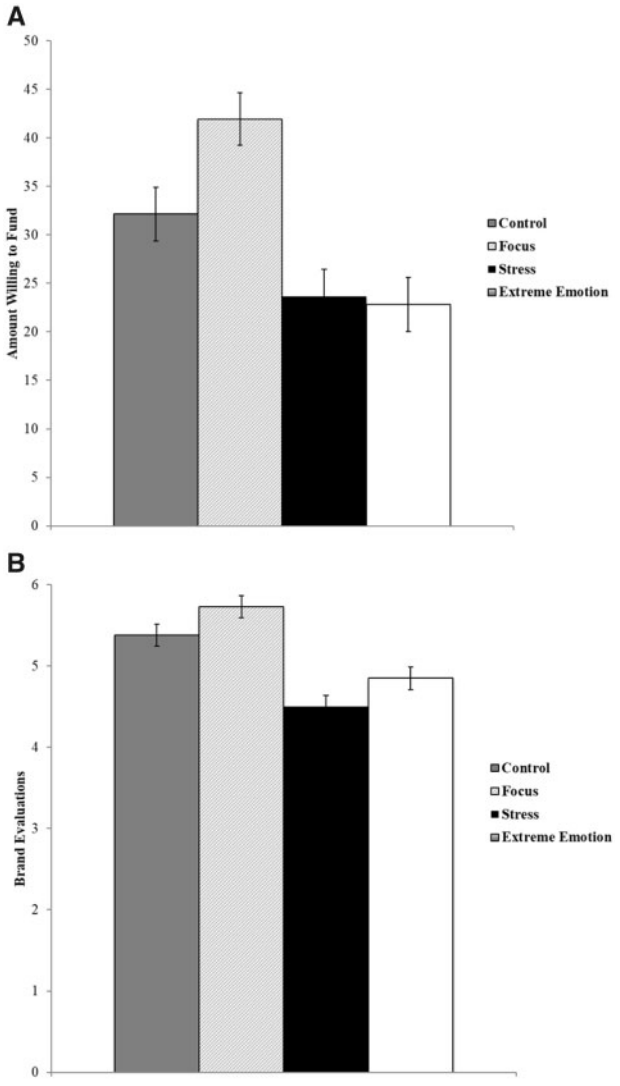
Underlying Mechanism Having experimentally validated our key findings from Studies 1 and 2 (H1–H3), we now examine the underlying mechanism. Perceived persuader competence is expected to mediate the relationship between the vocal tone and the outcome variables. We first ran a one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. focus) ANOVA with perceptions of persuader competence as the dependent variable. This revealed a significant main effect of vocal tone dimensions ($F(3, 409) = 52.444$, $p < .001$, $\eta^2 = .278$). As table 5 shows, participants inferred higher competence when the vocal tone exhibited signs of focus than in the control condition ($M_{focus} = 5.56$ vs. $M_{control} = 5.04$, $p = .006$). In contrast, participants reported lower, and marginally lower, perceived competence when the vocal tone exhibited signs of stress ($M_{stress} = 3.29$ vs. $M_{control} = 5.04$, $p = .001$) and signs of extreme emotion ($M_{extreme\ emotion} = 4.69$ vs. $M_{control} = 5.04$, $p = .066$), respectively.

We tested two regression models using vocal tone dimensions as the independent multicategorical variable, perceived persuader competence as the continuous mediator, and amount willing to fund and brand evaluations as the two continuous dependent variables (Model 4; Hayes 2017, with 5000 bootstraps). The results suggest that all the indirect effects (vocal tone dimensions → perceived persuader competence → amount willing to fund and brand evaluations) were significant and excluded zero for both outcome variables. In other words, compared to the control condition, the focused vocal tone resulted in greater perceptions of persuader competence, which in turn led participants to report higher amount willing to fund and brand evaluations ($b_{amount} = 2.74$, SE = 1.16; 95% CI = .75, 5.22; $b_{evaluations} = .22$, SE = .09; 95% CI = .07, .40). Conversely, compared to the control condition, vocal tones that sounded stressed ($b_{amount} = -9.73$, SE = 1.96; 95% CI = −13.81, −6.16; $b_{evaluations} = -.80$, SE = .13; 95% CI = −1.07, −.56) or extremely emotional ($b_{amount} = -2.03$, SE = 1.09; 95% CI = −4.25, −.01; $b_{evaluations} = -.17$, SE = .09; 95% CI = −.36, −.01) led to decreased perceived persuader competence, lowering amount willing to fund and brand evaluations. These results provide evidence for the mediating role of perceived persuader competence and support hypothesis 6.

Alternative Explanations A one-way (vocal tone dimension: control vs. extreme emotion vs. stress vs. focus) MANOVA with processing fluency and perceived trust as the dependent variables yielded no main effects of vocal tone dimensions (p's > .311). These results suggest that

FIGURE 5

(A) AMOUNT WILLING TO FUND AND (*B*) BRAND EVALUATIONS (STUDY 3)



processing fluency and perceived persuader trust did not account for our findings.

We also ran the multiple-mediation analyses (Model 4; Hayes 2017; with 5000 bootstraps). Perceived persuader competence, processing fluency, and perceived persuader trust entered simultaneously as mediators of the effect of vocal tone dimensions on amount willing to fund and brand evaluations. Our findings remained unchanged: perceived persuader competence emerged as a significant mediator on both amount willing to fund ($b_{\text{focus vs. control}} = 2.31$, SE $= 1.03$; 95% CI $= .55, 4.57$; $b_{\text{stress vs. control}} = -8.20$, SE $= 2.00$; 95% CI $= -12.23, -4.42$; $b_{\text{extreme emotion vs. control}} = -1.72$, SE $= .94$; 95% CI $= -3.74, -.02$) and brand evaluations ($b_{\text{focus vs. control}} = .21$, SE $= .08$; 95% CI $= .06, .38$; $b_{\text{stress vs. control}} = -.75$, SE $= .13$; 95% CI $= -1.04, -.50$; $b_{\text{extreme emotion vs. control}} = -.16$, SE $= .09$; 95% CI $= -.34, -.01$).

The indirect paths from the processing fluency mediator were nonsignificant on both amount willing to fund ($b_{\text{focus vs. control}} = .06$, SE $= .49$; 95% CI $= -.85, 1.19$; $b_{\text{stress vs. control}} = .76$, SE $= .56$; 95% CI $= -.11, 2.07$; $b_{\text{extreme emotion vs. control}} = .24$, SE $= .49$; 95% CI $= -.71, 1.33$) and brand evaluations ($b_{\text{focus vs. control}} = .001$, SE $= .01$; 95% CI $= -.02, .02$; $b_{\text{stress vs. control}} = .003$, SE $= .02$; 95% CI $= -.04, .04$; $b_{\text{extreme emotion vs. control}} = .001$, SE $= .01$; 95% CI $= -.03, .02$). Similarly, the indirect paths from the perceived persuader trust mediator were nonsignificant[5] on amount willing to fund ($b_{\text{focus vs. control}} = -.13$, SE $= .72$; 95% CI $= -1.47, 1.48$; $b_{\text{stress vs. control}} = -.78$, SE $= .78$; 95% CI $= -2.64, .49$; $b_{\text{extreme emotion vs. control}} = -.44$, SE $= .68$; 95% CI $= -2.01, .74$) and brand evaluations ($b_{\text{focus vs. control}} = -.004$, SE $= .02$; 95% CI $= -.05, .04$; $b_{\text{stress vs. control}} = -.02$, SE $= .02$; 95% CI $= -.08, .01$; $b_{\text{extreme emotion vs. control}} = -.01$, SE $= .02$; 95% CI $= -.06, .02$). As such, we fail to find evidence that processing fluency and perceived persuader trust mediate the effect of vocal tone dimensions on the outcome variables.

Discussion of Study 3

Study 3 demonstrates that a funding request is less successful when a persuader's voice shows greater signs of stress or extreme emotion, and more successful when a persuader's voice shows greater signs of focus. These findings establish a robust and differential role of vocal tones in a causal setting, supporting our proposed mediator (H6). Focused vocal tones (vs. control) generate greater perceptions of persuader competence, increasing willingness to fund, and improved brand evaluations; contrastingly, a stressed or extremely emotional vocal tone (vs. control) reduces perceived persuader competence, lowering the same factors. Study 3 also ruled out alternative explanations for the core effect[6]: vocal tone dimensions do not seem to foster different levels of processing fluency or persuader trust, and are therefore unlikely to drive our findings.

5 In line with prior research, perceived persuader trust significantly and positively predicted both amount willing to fund ($b_{\text{trust}} = .19$, $t(412) = 3.923$, $p = .001$) and brand evaluations ($b_{\text{trust}} = .21$, $t(412) = 4.419$, $p = .001$).

6 A further study, available as web appendix E, uses Kickstarter pitches selected for their high scores on the vocal tone measures (i.e., one shows high focus) and mitigates persuader warmth as an explanation for our core effects.

TABLE 6

CELL MEANS IN STUDY 4

	Condition		
	Control ($n = 105$)	Low pitch ($n = 104$)	High pitch ($n = 104$)
Amount willing to fund	36.24 (30.76)	46.15 (31.52)*	27.91 (27.32)*
Brand evaluations	5.31 (1.30)	5.76 (1.09)*	4.94 (1.62)*
Perceived competence	4.95 (1.40)	5.32 (1.21)**	4.47 (1.67)*
Signs of focus	4.85 (1.41)	5.31 (1.26)*	4.44 (1.59)*

Notes. Standard deviations in parentheses;
*a significant difference between a control and a vocal tone dimension at $p < .05$;
**a marginally significant difference at $p < .10$.

STUDY 4: HIGH VERSUS LOW PITCH ALTERATION TO VARY PERCEIVED FOCUS

Study 4 extends the prior studies in two key ways. First, given that sounding focused leads to favorable outcomes, a natural question arises: How does one sound focused online? Finding ways to do so greatly enhances the potential application of our work. An equally important second goal is to better connect our research to the literature on how vocal characteristics (e.g., high-pitched vs. low-pitched) impact inferences. Lower-pitched voices tend to be evaluated as belonging to more potent, strong, physically determined, competent, and successful speakers, raising the possibility that shifting pitch could alter perceptions of focus, and correspondingly, competence. Study 4 empirically investigates this prospect.

Method

Participants, Design, and Procedure We randomly assigned 313 MTurk volunteers (52% females, M_{age} = 30.9) to one of three conditions (voice pitch: control vs. low vs. high) in a single-factor between-subjects design in exchange for a small amount of financial compensation. This study was introduced as evaluating a new brand on Kickstarter—the COOLEST Cooler.

Manipulating Voice Pitch

We relied on prior work on vocal characteristics (Apple et al. 1979; Chattopadhyay et al. 2003; Cheng et al. 2016; Klofstad 2016; Oleszkiewicz et al. 2017; Tigue et al. 2012), which has demonstrated that voice pitches can be experimentally raised or lowered using voice and pitch shifting, modulating, and/or generating software. We used our pretested COOLEST Cooler control condition recording from study 3 as a baseline. Next, we altered the control condition recording to create two additional versions: a higher-pitched recording and a lower-pitched one. We used an open-source, free software package called Online Tone Generator.[7] This flexible voice and pitch shifter, modulator, and generator software relies on HTML5 Web Audio Application programming interface to alter a voice pitch without affecting its tempo (i.e., slowing down or speeding up). To create higher- versus lower-pitched recordings, the control condition pitch was shifted by +3 and −3 semitones, respectively (one semitone is equivalent to shifting by 5.946%).

After hearing the pitch, study participants responded to the dependent measures, amount willing to fund and brand evaluations. We then administered a voice pitch manipulation check and a measure of perceived focus. Afterward, participants rated perceived persuader competence before answering a short demographic and suspicion probe section.

Measures We measured amount willing to fund (α = .84), brand evaluations (α = .95), perceived persuader competence (α = .95), and signs of focus (α = .92), as in prior studies. The pitch manipulation check involved listeners indicating whether the persuader's voice sounded low-pitched or high-pitched (reverse-coded) on scales ranging from 1 (not at all) to 7 (very) (α = .75).

Manipulation Check A one-way (voice pitch: control vs. low vs. high) ANOVA with the voice pitch manipulation check items as the dependent variable revealed that our manipulation was successful ($F(2, 310) = 56.560$, p = .001, $\eta^2 = .267$). That is, participants in the low pitch condition reported higher scores on the voice pitch manipulation check compared to respondents in the high pitch condition ($M_{low\ pitch} = 4.79$ vs. $M_{high\ pitch} = 2.52$, p = .001), while the control condition participants scored in between ($M_{control} = 4.06$ vs. $M_{high\ pitch} = 2.52$, p = .001; $M_{control} = 4.06$ vs. $M_{low\ pitch} = 4.79$, p = .001).

Next, we examined whether shifting the voice pitch can help alter perceptions of concentration. In line with our expectations, the findings of a one-way (voice pitch: control vs. low vs. high) ANOVA with the signs of concentration as the outcome variable revealed a main effect parallel to that observed on the voice pitch check ($F(2, 310)$ =

7 http://onlinetonegenerator.com/pitch-shifter.html, Online Tone Generator, accessed August 19, 2020.

FIGURE 6

(A) AMOUNT WILLING TO FUND AND (*B*) BRAND EVALUATIONS (STUDY 4)

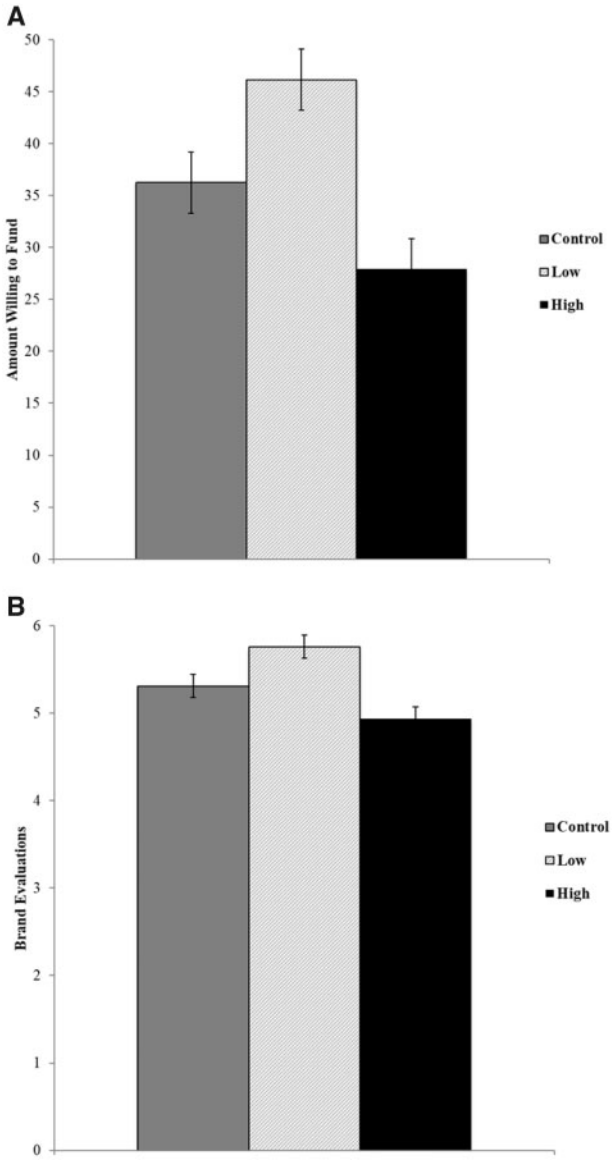
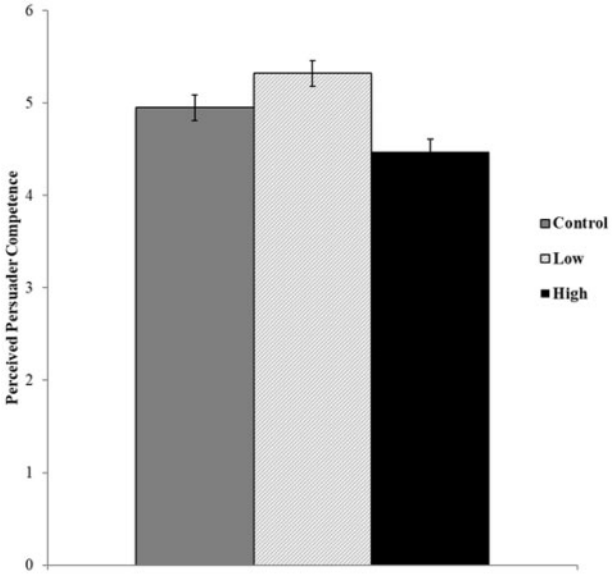


FIGURE 7

PERCEIVED PERSUADER COMPETENCE (STUDY 4)



9.668, $p = .001$, $\eta^2 = .059$). As table 6 shows, respondents in the low pitch condition reported higher scores on the signs of focus check compared to participants in the high pitch condition ($M_{\text{low pitch}} = 5.31$ vs. $M_{\text{high pitch}} = 4.44$, $p = .001$), with the control condition participants falling in the middle ($M_{\text{control}} = 4.85$ vs. $M_{\text{high pitch}} = 4.44$, $p = .042$; $M_{\text{control}} = 4.85$ vs. $M_{\text{low pitch}} = 5.31$, $p = .019$).

Amount Willing to Fund, Brand Evaluations, and Persuader Competence A one-way (voice pitch: control vs. low vs. high) MANOVA yielded a significant multivariate main effect of voice pitch on amount willing to fund, brand evaluations, and perceived persuader competence ($F(6, 616) = 4.94$, Wilk's $\lambda = .910$, $p = .001$, $\eta^2 = .046$). Follow-up univariate ANOVAs further examined the main effect on each variable. As predicted, a one-way (voice pitch: control vs. low vs. high) ANOVA with amount willing to fund as our dependent variable revealed a main effect of voice pitch ($F(2, 310) = 9.678$, $p < .001$, $\eta^2 = .059$). As figure 6*A* illustrates, the low-pitched voice elicited higher amount willing to fund than the control voice ($M_{\text{low pitch}} = 46.15$ vs. $M_{\text{control}} = 36.24$, $p = .017$), a boost of 27.4%. In contrast, the high-pitched voice generated lower amount willing to fund than the control voice ($M_{\text{high pitch}} = 27.91$ vs. $M_{\text{control}} = 36.24$, $p = .045$), a reduction of about 23%.

The findings for brand evaluations mirrored the aforementioned results (figure 6*B*). A one-way (voice pitch: control vs. low vs. high) follow-up univariate ANOVA with brand evaluations as the dependent variable yielded a main effect of voice pitch ($F(2, 310) = 9.656$, $p < .001$, $\eta^2 = .059$). The low-pitched voice resulted in enhanced brand evaluations compared to the control ($M_{\text{low pitch}} = 5.76$ vs. $M_{\text{control}} = 5.31$, $p = .016$) and high-pitched voices ($M_{\text{low pitch}} = 5.76$ vs. $M_{\text{high pitch}} = 4.94$, $p = .001$). Conversely, the high-pitched voice elicited reduced brand evaluations compared to those in the control voice condition ($M_{\text{high pitch}} = 4.94$ vs. $M_{\text{control}} = 5.31$, $p = .049$).

A one-way (voice pitch: control vs. low vs. high) ANOVA with perceptions of persuader competence as the

dependent variable showed a significant main effect of voice pitch, mirroring amount funded and brand evaluations ($F(2, 310) = 8.986$, $p < .001$, $\eta^2 = .055$; figure 7). Specifically, the low-pitched voice led to marginally greater inferred competence than the control condition ($M_{\text{low pitch}} = 5.32$ vs. $M_{\text{control}} = 4.95$, $p = .065$), whereas the high-pitched voice resulted in lowered inferred competence ($M_{\text{high pitch}} = 4.47$ vs. $M_{\text{control}} = 4.95$, $p = .018$).

Mediation If voice pitch can indeed change perceptions of focus and perceptions drive outcome variables, we should see a serial mediation of voice pitch's impact on the dependent variables (amount willing to fund and brand evaluations). To confirm this expectation, we tested two regression models using voice pitch as the independent multicategorical variable; signs of focus and perceived persuader competence as the proximal and distal, respectively, continuous mediators; and amount willing to fund and brand evaluations as the two continuous dependent variables (Model 6; Hayes 2017; with 5000 bootstraps).

All of the indirect effects (voice pitch → signs of focus → perceived persuader competence → amount willing to fund and brand evaluations) were significant and excluded zero for both amount willing to fund and brand evaluations. In other words, compared to the control condition, a low-pitched voice resulted in greater signs of focus ratings and subsequently increased perceptions of persuader competence, leading to higher willingness to fund and brand evaluations ($b_{\text{amount}} = 2.06$, SE = .98; 95% CI = .40, 4.31; $b_{\text{evaluations}} = .11$, SE = .05; 95% CI = .01, .23). Conversely, compared to the control, high-pitched voices led to lower signs of focus ratings, decreased perceived persuader competence, and lower amount willing to fund and brand evaluations ($b_{\text{amount}} = -1.79$, SE = 1.05; 95% CI = −4.01, −.03; $b_{\text{evaluations}} = -.09$, SE = .06; 95% CI = −.25, −.001). These results provide clear support for our theorizing.[8]

Discussion of Study 4

By using voice- and pitch-shifting software on a real Kickstarter campaign, we experimentally show that perceived persuader focus can be increased through the degree of pitch to obtain favorable funding and brand evaluation outcomes. Compared to both the control and high-pitched voice, the low-pitched voice increased signs of focus ratings, which in turn led to increases in perceived persuader competence, amount willing to fund, and brand evaluations. Thus, study 4 replicates the link between perceived persuader focus, competence, and funding outcomes, as well as providing preliminary evidence as to how perceptions of persuader focus can be increased. Importantly, these findings better theoretically connect the present work with the prior literature on vocal characteristics and consumers inferences.

Given that Studies 3 and 4 used a single female voice, we cannot dig into gender effects as much as this important topic deserves. It is unclear whether voice pitch would be a relatively simple main effect (i.e., lower pitch being better regardless of the speaker's gender) or whether some complex interaction with gender might be perceived. Further research would be most helpful.

GENERAL DISCUSSION AND FURTHER RESEARCH

We extend the consumer behavior literature on persuasion attempts by examining these attempts in a novel, nonlaboratory crowdfunding setting, and by conducting controlled experiments. In doing so, we move toward causal explanatory conclusions while showing the potential for real-world impact.

We echo prior findings that persuaders should carefully consider the signals they send beyond their words (Hall 1980; Nisbett and Wilson 1977). Our results suggest that to maximize success, persuaders should develop their pitches to a point where they are confident enough to avoid showing telltale signs of stress, which undermine perceptions of competence. Of course, this may require extensive practice for those who are not natural communicators.

Our results offer further practical guidelines. For instance, when making a request, persuaders should focus fully to help demonstrate their competence. To this end, we also show that low voice pitch generates perceptions of competence and persuasion success, thus bridging the consumer behavior literature on persuasion with the literature on vocal characteristics. Together, these findings contribute to SCM theory (Cuddy et al. 2008; Fiske et al. 2002, 2007). By uncovering that perceptions of competence underlie the differential impact of vocal tone on funding request success, we extend the SCM to a novel crowdfunding persuasion context.

Beyond our specific hypotheses, we contribute to the wider research by highlighting audio mining's potential in academic research. Fascinating new methods can quantify key elements of the human voice, promising benefits for numerous fields. While consumer research can take the lead in this area—gaining insight into a speaker's thoughts, identifying presenters' styles, and (as we do here) assessing persuasion effectiveness—any further research on both using and validating various audio mining techniques would be valuable.

The current research also speaks to online persuasion, a fruitful topic given the profusion of online interactions and additional venues in which to study persuasion (Grewal and Stephen 2019; Kupor and Tormala 2018). Indeed, we show the potential to use technology to change the signals

8 A further study, available as web appendix F, establishes additional robustness of study 4's findings by administering manipulation check items after the focal mediator as opposed to before.

sent (e.g., altering vocal pitch). While many online persuasion attempts may be quite different from those we examine, and offline persuasion has its own unique features, there are advantages to using Kickstarter for consumer research. All pitches are publicly accessible and follow a predictable format, which aids comparison between persuasion attempts. Moreover, crowdfunding sites contain a massive number of digitized observations. Data scraping can allow for countless observations with real-world relevance to be gained. We can even find out if artists have made prior requests for funds. As noted, our data shows that prior attempts are associated with less success; we assume that a prior attempt suggests a likely past failure and signals an artist's lack of appeal. That said, obviously there could be artists who gain experience and improve their pitches. The evolution of persuasion attempts would be a fascinating area to study. A final advantage of using Kickstarter data is that, because persuaders set their own targets, we know what they wanted, how they conducted their persuasion, and whether they ultimately achieved their desired outcome. Thus, a persuasion attempt's success can easily be observed.

At the same time, despite the benefits of using online data and audio mining techniques, we recognize the risks of drawing general conclusions from specific circumstances. Our investigation has considered one type of request, but there are multiple types of persuasion that people use, and are subject to, every day. An academic might seek to persuade students to work harder, a reviewer to look kindly upon a paper, or a child to do the washing up. Persuaders are also receivers. An advertiser might target the same academic with a new car, the college dean might have an exciting administrative assignment she wants filled, or the academic's spouse might angle for a preferred holiday destination. In short, persuasion is as diverse as human interaction. Further tests could be helpful in confirming which vocal tones matter and when.

We also note the technical problem that audio mining software requires audio clips of sufficient length and quality to generate effective readings. There are clips that can prove effective when using human subjects but that may still not be conductive to audio mining. Researchers should consider if they have the right audio clips for effective mining. The positive news is that audio-mining software is constantly progressing which should reduce such concerns over time.

Given our findings also imply that those seeking to persuade should limit their passion, lest it reduce perceptions of competence and hence their chances of funding, it would be valuable to better understand the boundaries of the effects. One can imagine circumstances where the display of extreme emotion might be beneficial to persuasion—for example, plausibly helping to signify an entrepreneur's passion. The possibility that conveying passion in a vocal tone can plausibly be sometimes a positive and sometimes a negative raises interesting questions about the role of context in better understanding the complex interaction between vocal tones and persuasion. Notably, different speakers in public life with very different styles are all able to persuade; it would be useful to better understand why (i.e., when and where exactly passion is an advantage).

Relatedly, an apt question might be this: For whom does passion work as a persuasion technique? Clearly, gender and other important social categorizations have a correlation with voice measures. The modern world is no stranger to prejudicial outcomes—both intentional and unintentional—and the presence of technology is no panacea to resolving long-standing social challenges in a world where the "default" is often seen as male (O'Neil 2016; Perez 2019). Given that women tend to have higher-pitched voices than men, this means any finding that lower-pitched voices are perceived more favorably is a source of concern in attempting to remove bias against women. Unfortunately, we do not here have the data to fully investigate such questions. Our first two studies use numerous pitches that vary on many dimensions beyond just gender, and our final two studies hold the gender of the speaker constant. As such, we hope that future work will help us better understand how people react to various voices and, critically, how we can mitigate any negative impacts.

This line of thought leads us to echo our general concerns about prejudice in decision making. Where people have intuitive reactions, and especially when they have no requirement to justify them, any reactions can be subject to unconscious bias. Sometimes these will be tied to an effect inextricably linked to the speaker—as we mentioned, there are clear differences in vocal tones and pitches between different groups of people. Furthermore, persuader and receiver perceptions might interact in a more subtle fashion. For instance, a member of a traditionally disadvantaged group might, due to past unfair experiences, feel more stress when presenting. As a result, any persuasion attempt made by this person may reflect the stress learned over a lifetime of experiencing prejudice, making him/her appear less competent to the receiver, who is unlikely to register the true reason for the stress. Instead the receiver makes false assumptions about a lack of competence. These types of complex interactions can help perpetuate social injustice even when the receiver has no such intention to do so. Accordingly, these possibilities represent a critical area for future research.

CONCLUSION

The vocal tones of persuaders can supply cues that are likely to impact the intuitions of receivers. Audio mining technology can measure these cues, and we apply it to identify the tones associated with successful Kickstarter

funding. Using two outcome measures in secondary data, we support our hypotheses that appeals are more successful when persuaders' vocal tones show greater focus, less stress, and less extreme emotion. With two controlled experiments, we then replicate our findings that vocal tone impacts funding decisions and also show that perceived competence mediates the relationship between vocal tone and persuasion attempt success. Our research introduces audio mining to the marketing literature and speaks to the value of combining secondary data and experiments when investigating persuasion.

DATA COLLECTION INFORMATION

The collection and coding of data were administered through the Kickstarter crowdfunding platform between April 2009 and December 2015 (Studies 1 and 2) and Amazon's Mechanical Turk in May 2019 (study 3) and Jan 2020 (study 4 and web appendix study). The third author scraped the data for the first two studies. The fourth author designed studies 3, 4, and web appendix study and carried out data collection and data analysis, with data and coding discussed on multiple occasions by all authors. The data are currently stored in a shared Dropbox folder.

REFERENCES

Aaker, Jennifer, Kathleen D. Vohs, and Cassie Mogilner (2010), "Nonprofits Are Seen as Warm and for-Profits as Competent: Firm Stereotypes Matter," *Journal of Consumer Research*, 37 (2), 224–37.

Addington, David W. (1968), "The Relationship of Selected Vocal Characteristics to Personality Perception," *Speech Monographs*, 35 (4), 492–503.

Agrawal, Ajay K., Christian Catalini, and Avi Goldfarb (2014), "Some Simple Economics of Crowdfunding," *Innovation Policy and the Economy*, 14 (1), 63–97.

Ampenberger, Andreas, Siegmund Staggl, and Wilfried Pohl (2017), "Attention Guidance, Perceived Brightness and Energy Demand in Retail Lighting," *Energy Procedia*, 111, 658–88.

Apple, William, Lynn A. Streeter, and Robert M. Krauss (1979), "Effects of Pitch and Speech Rate on Personal Attributions," *Journal of Personality and Social Psychology*, 37 (5), 715–27.

Asch, Solomon E. (1946), "Forming Impressions of Personality," *The Journal of Abnormal and Social Psychology*, 41 (3), 258–90.

Aucouturier, Jean-Julien, Petter Johansson, Lars Hall, Rodrigo Segnini, Lolita Mercadié, and Katsumi Watanabe (2016), "Covert Digital Manipulation of Vocal Emotion Alter Speakers' Emotional States in a Congruent Direction," *Proceedings of the National Academy of Sciences of United States of America*, 113 (4), 948–53.

Aylott, Russell and Vincent-Wayne Mitchell (1998), "An Exploratory Study of Grocery Shopping Stressors," *International Journal of Retail and Distribution Management*, 26 (9), 362–73.

Baker, Julie, Levy Michael, and Grewal Dhruv (1992), "An Experimental Approach to Making Retail Store Environmental Decisions," *Journal of Retailing*, 68 (4), 445–60.

Baker, Stacey Menzel, Jonna Holland, and Carol Kaufman-Scarborough (2007), "How Consumers with Disabilities Perceive 'Welcome' in Retail Servicescapes: A Critical Incident Study," *Journal of Services Marketing*, 21 (3), 160–73.

Batra, Rajeev and Douglas M. Stayman (1990), "The Role of Mood in Advertising Effectiveness," *Journal of Consumer Research*, 17 (2), 203–14.

Bayus, Barry L. (2013), "Crowdsourcing New Product Ideas Over Time: An Analysis of the Dell IdeaStorm Community," *Management Science*, 59 (1), 226–44.

Berry, Diane S. (1992), "Vocal Types and Stereotypes: Joint Effects of Vocal Attractiveness and Vocal Maturity on Person Perception," *Journal of Nonverbal Behavior*, 16 (1), 41–54.

Berry, Diane S., Jane S. Hansen, Julie C. Landry-Pester, and Jo A. Meier (1994), "Vocal Determinants of First Impressions of Young Children," *Journal of Nonverbal Behavior*, 18 (3), 187–97.

Boidron, Laurent, Karim Boudenia, Christophe Avena, Jean-Michel Boucheix, and Jean-Julien Aucouturier (2016), "Emergency Medical Triage Decisions Are Swayed by Computer-Manipulated Cues of Physical Dominance in Caller's Voice," *Scientific Reports*, 6 (30219), 1–7.

Brown, Bruce L., William J. Strong, and Alvin C. Rencher (1975), "Acoustic Determinants of Perceptions of Personality from Speech," *International Journal of the Sociology of Language*, 1975 (6), 11–32.

Carlo, Gustavo, Maria Vicenta Mestre, Meredith M. McGinley, Paula Samper, Ana Tur, and Deanna Sandman (2012), "The Interplay of Emotional Instability, Empathy, and Coping on Prosocial and Aggressive Behaviors," *Personality and Individual Differences*, 53 (5), 675–80.

Chattopadhyay, Amitava, Darren W. Dahl, Robin J.B. 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.

Chaudhuri, Arjun and Morris B. Holbrook (2001), "The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty," *Journal of Marketing*, 65 (2), 81–93.

Chen, Xiao-Ping, Xin Yao, and Suresh Kotha (2009), "Entrepreneur Passion and Preparedness in Business Plan Presentations: A Persuasion Analysis of Venture Capitalists' Funding Decisions," *Academy of Management Journal*, 52 (1), 199–214.

Cheng, Joey T., Jessica L. Tracy, Simon Ho, and Joseph Henrich (2016), "Listen, Follow Me: Dynamic Vocal Signals of Dominance Predict Emergent Social Rank in Humans," *Journal of Experimental Psychology: General*, 145 (5), 536–47.

Cholakova, Magdalena and Bart Clarysse (2015), "Does the Possibility to Make Equity Investments in Crowdfunding Projects Crowd out Reward Based Investments?" *Entrepreneurship Theory and Practice*, 39 (1), 145–72.

Collins, Sarah A. and Caroline Missing (2003), "Vocal and Visual Attractiveness Are Related in Women," *Animal Behaviour*, 65 (5), 997–1004.

Colombo, Massimo G., Chiara Franzoni, and Cristina Rossi Lamastra (2015), "Internal Social Capital and the Attraction of Early Contributions in Crowdfunding," *Entrepreneurship Theory and Practice*, 39 (1), 75–100.

Csikszentmihalyi, Mihaly. (1977), *Beyond Boredom and Anxiety*, San Francisco, CA: Jossey-Bass.

Cuddy, Amy J. C., Susan T. Fiske, and Peter Glick (2008), "Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map," *Advances in Experimental Social Psychology*, 40, 61–149.

Cudmore, Aileen and Darina M. Slattery (2019), "An Analysis of Physical and Rhetorical Characteristics of Videos Used to Promote Technology Projects, on the Kickstarter Crowdfunding Platform," *Technical Communication*, 66 (4), 319–46.

Custers, P. J. M., Yvonne de Kort, W. A. IJsselsteijn, and M. E. de Kruiff (2010), "Lighting in Retail Environments: Atmosphere Perception in the Real World," *Lighting Research and Technology*, 42 (3), 331–43.

Dillard, James Price and Lijiang Shen, ed. (2013), *The Sage Handbook of Persuasion*, Thousand Oaks, CA: Sage.

Dubois, David, Derek D. Rucker, and Adam D. Galinsky (2016), "Dynamics of Communicator and Audience Power: The Persuasiveness of Competence versus Warmth," *Journal of Consumer Research*, 43 (1), 68–85.

Elliot, Andrew J. and Judith M. Harackiewicz (1994), "Goal Setting, Achievement Orientation, and Intrinsic Motivation: A Mediational Analysis," *Journal of Personality and Social Psychology*, 66 (5), 968–80.

Fan, Tingting, Leilei Gao, and Yael Steinhart (2020), "The Small Predicts Large Effect in Crowdfunding," *Journal of Consumer Research*, 47 (4), 544–65.

Fiske, Susan T., Amy J. C. Cuddy, and Peter Glick (2007), "Universal Dimensions of Social Cognition: Warmth and Competence," *Trends in Cognitive Sciences*, 11 (2), 77–83.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu (2002), "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition," *Journal of Personality and Social Psychology*, 82 (6), 878–902.

Fiske, Susan T., Chris Malone, and Nicolas Kervyn (2012), "Brands as Intentional Agents: Our Response to Commentaries," *Journal of Consumer Psychology*, 22 (2), 205–7.

Fiske, Susan T. and Steven L. Neuberg (1990), "A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation," *Advances in Experimental Social Psychology*, 23, 1–74.

Giddens, Cheryl L., Kirk W. Barron, Jennifer Byrd-Craven, Keith F. Clark, and A. Scott Winter (2013), "Vocal Indices of Stress: A Review," *Journal of Voice*, 27 (3), e21–9.

Grant, Sharon and Kate Ferris (2012), "Identifying Sources of Occupational Stress in Entrepreneurs for Measurement," *International Journal of Entrepreneurial Venturing*, 4 (4), 351–73.

Grewal, Lauren and Andrew T. Stephen (2019), "In Mobile We Trust: The Effects of Mobile versus Nonmobile Reviews on Consumer Purchase Intentions," *Journal of Marketing Research*, 56 (5), 791–808.

Guo, Wenbo, Detmar Straub, Xiao Han, and Pengzhu Zhang (2017), "Understanding Vendor Preference in the Crowdsourcing Marketplace: The Influence of Vendor-Task Fit and Swift Trust," *Journal of Electronic Commerce Research*, 18 (1), 1–17.

Hall, Judith A. (1980), "Voice Tone and Persuasion," *Journal of Personality and Social Psychology*, 38 (6), 924–34.

Harker, LeeAnne and Dacher Keltner (2001), "Expressions of Positive Emotion in Women's College Yearbook Pictures and Their Relationship to Personality and Life Outcomes across Adulthood," *Journal of Personality and Social Psychology*, 80 (1), 112–24.

Hayes, Andrew F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, New York, NY: Guildford Press.

Hildebrand, Diogo, Yoshiko DeMotta, Sankar Sen, and Ana Valenzuela (2017), "Consumer Responses to Corporate Social Responsibility (CSR) Contribution Type," *Journal of Consumer Research*, 44 (4), 738–58.

Hobson, Jessen L., William J. Mayew, and Mohan Venkatachalam (2012), "Analyzing Speech to Detect Financial Misreporting," *Journal of Accounting Research*, 50 (2), 349–92.

Hollien, Harry (1980), "Vocal Indicators of Psychological Stress," *Annals of the New York Academy of Sciences*, 347 (1 Forensic Psyc), 47–71.

Hu, Ming, Xi Li, and Mengze Shi (2015), "Product and Pricing Decisions in Crowdfunding," *Marketing Science*, 34 (3), 331–45.

Hu, Minqing and Bing Liu, (2004), "Mining and Summarizing Customer Reviews," Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 168–77.

Jiang, Zhenhui and Izak Benbasat (2007), "The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding," *MIS Quarterly*, 31 (3), 475–500.

Kervyn, Nicolas, Susan T. Fiske, and Chris Malone (2012), "Brands as Intentional Agents Framework: How Perceived Intentions and Ability Can Map Brand Perception," *Journal of Consumer Psychology*, 22 (2), 166–76.

Khamitov, Mansur, Xin (Shane) Wang, and Matthew Thomson (2019), "How Well Do Consumer-Brand Relationships Drive Customer Brand Loyalty? Generalizations from a Meta-Analysis of Brand Relationship Elasticities," *Journal of Consumer Research*, 46 (3), 435–59.

Kickstarter (2020), "Stats," https://www.kickstarter.com/help/stats.

Kimble, Charles E. and Steven D. Seidel (1991), "Vocal Signs of Confidence," *Journal of Nonverbal Behavior*, 15 (2), 99–105.

Klofstad, Casey A. (2016), "Candidate Voice Pitch Influences Election Outcomes," *Political Psychology*, 37 (5), 725–38.

Klofstad, Casey A., Rindy C. Anderson, and Stephen Nowicki (2015), "Perceptions of Competence, Strength, and Age Influence Voters to Select Leaders with Lower-Pitched Voices," *PLoS One*, 10 (8), e0133779.

Klofstad, Casey A., Rindy C. Anderson, and Susan Peters (2012), "Sounds Like a Winner: Voice Pitch Influences Perception of Leadership Capacity in Both Men and Women," *Proceedings of the Royal Society B: Biological Sciences*, 279 (1738), 2698–704.

Koch, Jascha-Alexander and Qian Cheng (2016), "The Role of Qualitative Success Factors in the Analysis of Crowdfunding Success: Evidence from Kickstarter," PACIS 2016 Proceedings, Chiayi, Taiwan, 242.

Koch, Jascha-Alexander and Michael Siering (2019), "The Recipe of Successful Crowdfunding Campaigns," *Electronic Markets*, 29 (4), 661–79.

Korzynski, Pawel, Michael Haenlein, and Mika Rautiainen (2021), "Impression Management Techniques in Crowdfunding: An Analysis of Kickstarter Videos Using Artificial Intelligence," *European Management Journal*, in press.

Krishna, Aradhna (2012), "An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior," *Journal of Consumer Psychology*, 22 (3), 332–51.

Krishna, Aradhna and Norbert Schwarz (2014), "Sensory Marketing, Embodiment, and Grounded Cognition: A Review and Introduction," *Journal of Consumer Psychology*, 24 (2), 159–68.

Kupor, Daniella and Zakary Tormala (2018), "When Moderation Fosters Persuasion: The Persuasive Power of Deviatory Reviews," *Journal of Consumer Research*, 45 (3), 490–510.

Lee, Angela Y. and Jennifer L. Aaker (2004), "Bringing the Frame into Focus: The Influence of Regulatory Fit on Processing Fluency and Persuasion," *Journal of Personality and Social Psychology*, 86 (2), 205–18.

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.

Li, Xueni (Shirley), Kimmy Wa Chan, and Sara Kim (2019), "Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters," *Journal of Consumer Research*, 45 (5), 973–87.

Lin, Mingfeng and Siva Viswanathan (2016), "Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market," *Management Science*, 62 (5), 1393–31.

Liu, Bing, Minqing Hu, and Junsheng Cheng (2005), "Opinion Observer: Analyzing and Comparing Opinions on the Web," Chiba, Japan: Proceedings of the 14th International World Wide Web Conference.

Mayew, William J. and Mohan Venkatachalam (2012), "The Power of Voice: Managerial Affective States and Future Firm Performance," *The Journal of Finance*, 67 (1), 1–43.

Meuter, Matthew L., Amy L. Ostrom, Mary Jo Bitner, and Robert Roundtree (2003), "The Influence of Technology Anxiety on Consumer Use and Experiences with Self-Service Technologies" *Journal of Business Research*, 56 (11), 899–906.

Mick, David Glen and Susan, Fournier (1998), "Paradoxes of Technology: Consumer Cognizance, Emotions, and Coping Strategies," *Journal of Consumer Research*, 25 (2), 123–43.

Mohan, Geetha, Bharadhwaj Sivakumaran, and Piyush Sharma (2013), "Impact of Store Environment on Impulse Buying Behavior," *European Journal of Marketing*, 47 (10), 1711–32.

Mollick, Ethan (2014), "The Dynamics of Crowdfunding: An Exploratory Study," *Journal of Business Venturing*, 29 (1), 1–16.

Mollick, Ethan and Ramana Nanda (2016), "Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts," *Management Science*, 62 (6), 1533–41.

Morgan, Robert M. and Shelby D. Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing*, 58 (3), 20–38.

Nemesysco (2015), *Call Center Quality Assurance*, Kadima, Israel.

—— (2016), The QA5 Call Center Software Technology, http://qa5system.com/qa5_technology.html

Nisbett, Richard E. and Timothy D. Wilson (1977), "The Halo Effect: Evidence for Unconscious Alteration of Judgments," *Journal of Personality and Social Psychology*, 35 (4), 250–6.

Nurttila, Suvi, Elina Ketonen, and Kirsti Lonka (2015), "Sense of Competence and Optimism as Resources to Promote Academic Engagement," *Procedia – Social and Behavioral Sciences*, 171, 1017–26.

O'Neil, Cathy (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York City, NY: Broadway Books.

Oleszkiewicz, Anna, Katarzyna Pisanski, Kinga Lachowicz-Tabaczek, and Agnieszka Sorokowska (2017), "Voice-Based Assessments of Trustworthiness, Competence, and Warmth in Blind and Sighted Adults," *Psychonomic Bulletin and Review*, 24 (3), 856–62.

Perez, Caroline Criado. (2019), *Invisible Women: Exposing Data Bias in a World Designed for Men*, New York City, NY: Random House.

Proulx, Michael J. and Howard E. Egeth (2007), "Biased Competition and Visual Search: The Role of Luminance and Size Contrast," *Psychological Research*, 72 (1), 106–13.

Rachman, Laura, Marco Liuni, Pablo Arias, Andreas Lind, Petter Johansson, Lars Hall, Daniel Richardson, Katsumi Watanabe, Stéphanie Dubal, and Jean-Julien Aucouturier (2018), "DAVID: An Open-Source Platform for Real-Time Transformation of Infra-Segmental Emotional Cues in Running Speech," *Behavior Research Methods*, 50 (1), 323–43.

Rafaeli, Anat and Robert I. Sutton (1989), "The Expression of Emotion in Organizational Life," *Research in Organizational Behavior*, 11 (1), 1–42.

Ray, George B. (1986), "Vocally Cued Personality Prototypes: An Implicit Personality Theory Approach," *Communication Monographs*, 53 (3), 266–76.

Reber, Rolf, Piotr Winkielman, and Norbert Schwarz (1998), "Effects of Perceptual Fluency on Affective Judgments," *Psychological Science*, 9 (1), 45–8.

Rockwell, Patricia (1996), "The Effects of Vocal Variation on Listener Recall," *Journal of Psycholinguistic Research*, 25 (3), 431–41.

Scherer, Klaus R., Harvey London, and Jared J. Wolf (1973), "The Voice of Confidence: Paralinguistic Cues and Audience Evaluation," *Journal of Research in Personality*, 7 (1), 31–44.

Shields, Stephanie A. (2005), "The Politics of Emotion in Everyday Life: 'Appropriate' Emotion and Claims on Identity," *Review of General Psychology*, 9 (1), 3–15.

Street, Richard L. and Robert M. Brady (1982), "Speech Rate Acceptance Ranges as a Function of Evaluative Domain, Listener Speech Rate, and Communication Context," *Communication Monographs*, 49 (4), 290–308.

Street, Richard L., Robert M. Brady, and William B. Putman (1983), "The Influence of Speech Rate Stereotypes and Rate Similarity or Listeners' Evaluations of Speakers," *Journal of Language and Social Psychology*, 2 (1), 37–56.

Streeter, Lynn A., Robert M. Krauss, Valerie Geller, Christopher Olson, and William Apple (1977), "Pitch Changes During Attempted Deception," *Journal of Personality and Social Psychology*, 35 (5), 345–50.

Summers, Teresa A. and Paulette R. Hebert (2001), "Shedding Some Light on Store Atmospherics: Influence of Illumination on Consumer Behavior," *Journal of Business Research*, 54 (2), 145–50.

Swan, John E., Michael R. Bowers, and Lynne D. Richardson (1999), "Customer Trust in the Salesperson: An Integrative Review and Meta-Analysis of the Empirical Literature," *Journal of Business Research*, 44 (2), 93–107.

Swan John, I. Fred Trawick, David R. Rink, and Jenny J. Roberts (1988), "Measuring Dimensions of Purchaser Trust of Industrial Salespeople," *Journal of Personal Selling and Sales Management*, 8 (1), 1–10.

Tigue, Cara C., Diana J. Borak, Jillian J. M. O'Connor, Charles Schandl, and David R. Feinberg (2012), "Voice Pitch Influences Voting Behavior," *Evolution and Human Behavior*, 33 (3), 210–6.

Van Zant, Alex B. and Jonah Berger (2020), "How the Voice Persuades," *Journal of Personality and Social Psychology*, 118 (4), 661–82.

Wang, Ze, Mao Huifang, Yexin Jessica Li, and Liu Fan (2017), "Smile Big or Not? Effects of Smile Intensity on Perceptions of Warmth and Competence," *Journal of Consumer Research*, 43 (5), 787–805.

Wiener, Hillary J. D. and Tanya L. Chartrand (2014), "The Effect of Voice Quality on Ad Efficacy," *Psychology & Marketing*, 31 (7), 509–17.

Wincent, Joakim, Daniel Örtqvist, and Mateja Drnovsek (2008), "The Entrepreneur's Role Stressors and Proclivity for a Venture Withdrawal," *Scandinavian Journal of Management*, 24 (3), 232–46.

Winkielman, Piotr and John T. Cacioppo (2001), "Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Elicits Positive Affect," *Journal of Personality and Social Psychology*, 81 (6), 989–1000.

Yang, Linyun W. and Pankaj Aggarwal (2019), "No Small Matter: How Company Size Affects Consumer Expectations and Evaluations," *Journal of Consumer Research*, 45 (6), 1369–84.

Zhang, Lanfang, Junfeng Zhang, Yangzexi Liu, and Jingqiu Guo (2017), "The Impact of Over-Bright Highway Billboards on Driving Behavior," in *Paper Presented at Third International Conference on Vehicle Technology and Intelligent Transport Systems*, ed. Oleg Gusikhin, Porto, Portugal: VEHITS.

Zhang, Shunyuan, Lee Dokyun, Singh Param, and Srinivasan. Kannan (2019), "How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics," Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2976021.

Zhou, Huiquan and Shihua Ye (2019), "Legitimacy, Worthiness, and Social Network: An Empirical Study of the Key Factors Influencing Crowdfunding Outcomes for Nonprofit Projects," *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 30 (4), 849–64.