Designing Persuasive Crowdfunding Videos

Hannah Chang
Singapore Management University
Anirban Mukherjee
Cornell University
Amitava Chattopadhyay
INSEAD

Acknowledgements:
The research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2 grant, MOE2018-T2-1-181, to the first author. The authors thank Matthew Blower, Joyce H.K. Chua, and Alina Wong for their research assistance.

Cite as:
Chang Hannah, Mukherjee Anirban, Chattopadhyay Amitava (2022), Designing Persuasive Crowdfunding Videos. *Proceedings of the European Marketing Academy*, 51st, (107605)

Paper from the 51st Annual EMAC Conference, Budapest, May 24-27, 2022

Designing Persuasive Crowdfunding Videos

Abstract:

Voiceover narration is a production technique commonly used in reward-based crowdfunding videos. We posit that in these videos, hearing more narrator voices describing the crowdfunding product can systematically influence consumers' attention and processing of the message, thereby facilitating persuasion. We employed a multi-method approach, including experimentation, natural language processing, text mining, and machine learning. Results across four studies—including real-world datasets and controlled experiments—show that the effect (1) has consequential, economic implications in a real-world marketplace, (2) is more pronounced when the message is easier to comprehend, and (3) is more pronounced when consumers have the capacity to process the narrated message. Substantive and theoretical implications are discussed.

Keywords: crowdfunding, video, persuasion

Track: Consumer Behavior

1. Introduction of Paper

Consider two videos: In one video, multiple narrator voices sequentially describe the features of a new product innovation. In another video, a single narrator voice conveys the same product message. In viewing such videos, does the number of voices narrating a message affect consumers' processing of the message and subsequent behavior? If so, is the effect disruptive or facilitative? In this research, we examine these questions in the context of crowdfunding videos, which have become increasingly prevalent and important in consumer decision making.

We posit and show that in crowdfunding videos, consumers who hear a persuasive message voiced by multiple narrators (vs. one narrator) would be more receptive to the new product, due to increased consumer attention and processing. Our rationale is as follows: As consumers' cognitive resources are constrained, they tend to deliberate on a subset of features which capture their attention (Lynch & Srull, 1982). Prior research has shown that a change in voice can involuntarily capture attention, even when other tasks are competing for attention (Cherry, 1953). We propose that when a new narrator's voice carries on a product message, there should be a momentary increase in consumers' attention and processing of the next piece of spoken message that might not have been processed otherwise. The enhanced processing of the message would increase the persuasive appeal of the product. We term this phenomenon—the voice numerosity effect.

We test our predictions in two studies (plus validation and replication studies), using a large-scale, real-world dataset (e.g., in the Kickstarter dataset, with more than 11,000 crowdfunding videos and over 3.6 million customer transactions) and a controlled experiment (with over 2,200 participants). Study 1 examines the hypothesized effect by using a dataset we collected from Kickstarter, a leading crowdfunding platform for innovations, with consequential dependent measures. We processed the video data using machine learning, text mining, and natural language processing; we then linked the video and project characteristics to consequential dependent measures in the crowdfunding context. The next study is a controlled experiment to further test our conceptualization underlying the effect—that it is driven by enhanced cognitive processing—by directly varying consumers' processing resources. Due to space constraints, we will only focus on these studies below and briefly discuss the validation and replication studies.

2. A Large-scale Dataset: Reward-based Crowdfunding Videos

We situated this empirical investigation of voice numerosity in the context of online crowdfunding, where product videos are commonly used to communicate innovations to potential consumers. We collected data from Kickstarter, which allows us to explore the hypothesized effect in a real-world marketplace (Dhanani & Mukherjee, 2019). We also tested our conceptualization that this effect relates to consumers' cognitive processing. Building on prior studies that show faster speech rate disrupts listeners' cognitive processing (e.g., Moore, Hausknecht, and Thamodaran, 1986), we predict having more narrator voices to communicate a product message can enhance its persuasiveness. This improves project outcomes, when the message is spoken at a slower rate (when speech is easier to process) but not when the message is spoken at a faster rate (when speech is more difficult to process).

2.1 Research setting and data preparation

We downloaded project webpages and videos from Kickstarter between July 1, 2017 and December 31, 2019, in 31 categories under the three largest supra-categories: product design, technology, and games. The resulting sample has 11,801 US-based projects, with over 3.6 million customer transactions. On Kickstarter, entrepreneurs pre-sell new products to prospective customers, thereby raising the capital to manufacture and launch these products. Customers browse the project webpage, which includes a video and its funding goal, to determine if they would pre-purchase the product. They pledge funds in return for the product to be shipped to them after it is manufactured.

We examined three consequential dependent variables relevant to this context: (1) the amount of funding pledged to support a project, (2) the number of backers, and (3) whether the project is successfully funded (Fan, Gao, and, Steinhart, 2020; Younkin & Kuppuswamy, 2018). The key independent variables are the number-of-voices and speech rate.

Figure 1 provides an overview of how we processed the raw (video) data and the measures derived. To parse the videos, we utilized a multimedia library to decouple the audio from the images. To derive our focal and control variables, we applied a deep-learning-based speech-recognition model to the audio track. It operates on the waveform to extract the data features needed to deduce "who said what" (Makino et al., 2019), enabling us to measure the number of unique narrator voices in each audio. We utilized various machine learning methods to measure four sets of control variables relating to persuasion, following prior research. These are controls to account for (1) audial features of the video (e.g., Goldberg, Chattopadhyay,

Gorn, and Rosenblatt, 1993), (2) linguistic features of the spoken verbal content (Pennebaker et al., 2015), (3) project-level characteristics (Li et al., 2019) and (4) visual elements (Liu et al., 2018; Li et al, 2019).

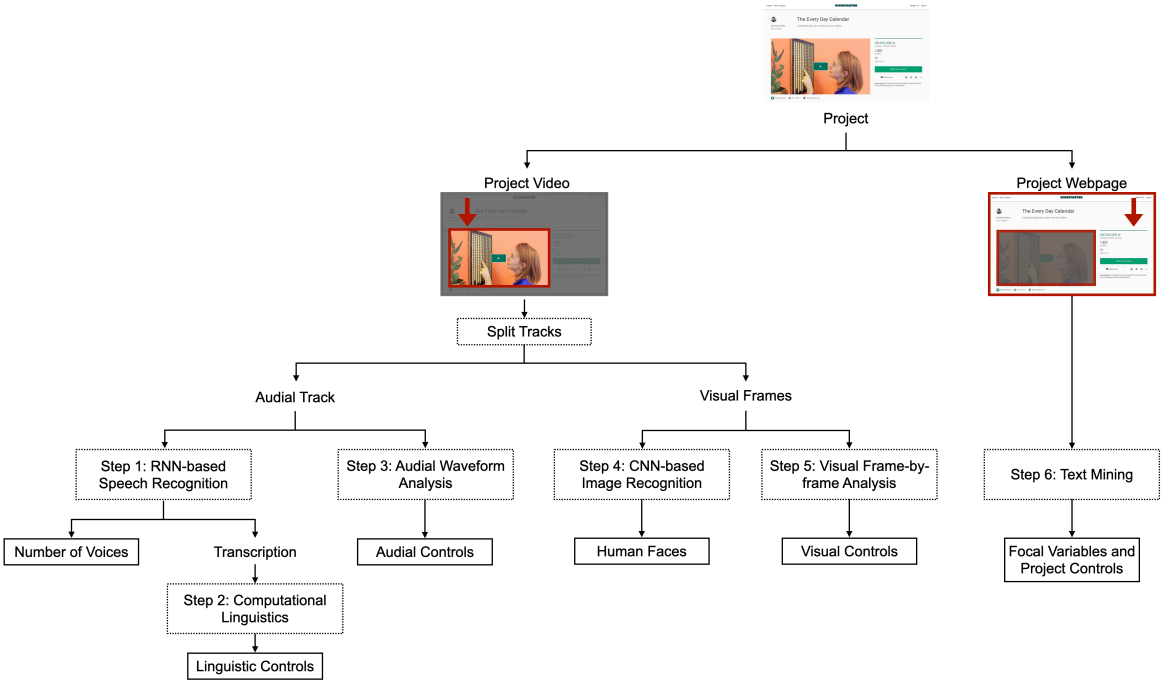

Figure 1. Overview of Variable Construction in Crowdfunding Dataset

2.2 Empirical analysis and results

We estimate a Type-I Tobit model for the amount of funding pledged (as it is left-censored at 0):

$$\begin{aligned}\text{pledged}_p^* = \; & \alpha_0 \\ & + \alpha_1 \times \text{num\_voices}_p + \alpha_2 \times \text{rate}_p + \alpha_3 \times \text{num\_voices}_p \times \text{rate}_p \\ & + \beta_1 \times \text{audial entropy}_p + \beta_2 \times \text{duration}_p + \beta_3 \times \text{energy}_p + \beta_4 \times \text{spectral\_centroid}_p \\ & + \beta_5 \times \text{spectral\_entropy}_p + \beta_6 \times \text{volume}_p + \beta_7 \times \text{ZCR}_p \\ & + \eta_1 \times \text{analytic}_p + \eta_2 \times \text{authentic}_p + \eta_3 \times \text{clout}_p + \eta_4 \times \text{tone}_p + \eta_5 \times \text{num\_words}_p \\ & + \delta_1 \times \text{creator\_experience}_p + \delta_2 \times \text{funding\_goal}_p \\ & + \delta_3 \times \text{menu\_length}_p + \delta_4 \times \text{menu\_price}_p + \delta_5 \times \text{proj\_duration}_p \\ & + \theta_1 \times \text{faces}_p + \theta_2 \times \text{scenes}_p + \theta_3 \times \text{visual\_variation}_p \\ & + \sum_{i=2}^{12} \gamma_{mi}\, \text{month}_{pi} + \sum_{j=2018}^{2019} \gamma_{yj}\, \text{year}_{pj} + \sum_{k=2}^{32} \gamma_{ck}\, \text{category}_{pk} + \varepsilon_{p,}\end{aligned} \tag{1}$$

where pledged_p^* is observed if $\text{pledged}_p^* \geq 0$ and 0 otherwise. α_1 measures the impact of the number of voices on the pledged amount in project p. Control variables for audial, linguistic, project, and visual elements correspond to βs, ηs, δs, and θs, respectively. The fixed effects $\left\{ \{\gamma_{mi}\}_{i=2}^{12}, \{\gamma_{yj}\}_{j=2018}^{2019}, \{\gamma_{ck}\}_{k=2}^{21} \right\}$ account for seasonal, annual, and category-specific differences across videos. ε_p is the error term.

We estimated analogous models to that of equation (1) for the other key dependent measures. We specified a Type-I Tobit model for number of backers (as it is also left-censored at 0) and a Probit regression model for project success (as it has binary outcome), where we observe $success_p$ if $pledged_p \geq goal_p$.

Table 1 presents our results. Across all three dependent variables, we found that having more voices narrating the project message significantly improves project outcomes (all $\alpha_1 > 0$, all *p*s < .01). This effect is both statistically significant and economically important—having an additional speaker, ceteris paribus, is associated with raising $12,795 in additional funds, getting the support of 118 additional backers, and a 1.6% increase in the probability of project success. Moreover, we found that the effect is consistently moderated by the rate at which the spoken content was delivered (all $\alpha_3 < 0$, all *p*s < .05), suggesting that cognitive processing is necessary for the effect of voice numerosity. To further investigate the nature of this interaction, we examined the marginal effect of the number of narrating voices on the pledged amount, number of backers, and project success. Across all three project outcomes, results consistently showed that the benefit of an additional voice is higher for easier-to-comprehend videos (slower rate; e.g., one spoken word per second) than for more complex videos (faster rate; e.g., three spoken words per second).

The findings are replicated across all three dependent measures that are of importance to the crowdfunding context. They are also replicated across the three largest supra-categories on Kickstarter (spanning 31 categories)—Design, Technology, and Games—which combine to almost 70% of all funds raised on the platform.

	Dependent variables: (1) Pledged amount (in USD), (2) Number of backers, (3) Project success		
	(1)	(2)	(3)
(Intercept)	-1,426,847.00*	-12,566.02	-13.01
	(722,293.40)	(6,751.72)	(14.29)
Number of voices	12,794.71***	117.85***	.25***
	(3,708.29)	(34.88)	(.07)
Rate	4,381.49	58.98*	.13*
	(2,761.26)	(25.89)	(.05)
Number of voices × Rate	-3,909.07*	-36.65*	-.09**
	(1,577.97)	(14.83)	(.03)
Audial controls			
Audial entropy	38,972.28***	394.61***	.46**
	(7,203.95)	(67.58)	(.14)
Duration	-40.06	-.36	-.001
	(26.45)	(.25)	(.001)
Energy	-7.70	-.08	-.0002

	Dependent variables: (1) Pledged amount (in USD), (2) Number of backers, (3) Project success		
	(1)	(2)	(3)
	(4.94)	(.05)	(.0001)
Spectral centroid	2,675,876.00	23,031.22	23.54
	(1,445,345.00)	(13,510.69)	(28.60)
Spectral entropy	11,352.80	94.56	.19
	(6,316.95)	(59.29)	(.12)
Volume	1,176.34***	11.58***	.03***
	(349.02)	(3.27)	(.01)
ZCR	3.05	.03	-0.0000
	(5.54)	(.05)	(.0001)
Linguistic controls			
Analytical thinking	114.19***	.92**	.001
	(31.68)	(.30)	(.001)
Authenticity	37.50	.32	.0003
	(34.56)	(.32)	(.001)
Clout	139.90*	1.52**	-.001
	(57.93)	(.54)	(.001)
Emotional tone	-58.76*	-.87***	.001**
	(26.80)	(.25)	(.001)
Number of words	27.56**	.20*	.001*
	(10.65)	(.10)	(.0002)
Project controls			
Creator experience	3,590.89***	40.64***	.10***
	(206.64)	(1.93)	(.01)
Funding goal	.30***	.002***	NA[1]
	(.02)	(.0001)	
Menu length	1,771.40***	16.03***	.04***
	(154.43)	(1.45)	(.003)
Menu price	-.26	-.10***	-.0002***
	(1.62)	(.02)	(0.0000)
Project duration	-117.18	-1.83**	-.02***
	(67.10)	(.63)	(.001)
Visual controls			
Faces	-394.69	-2.22	-.02**
	(342.01)	(3.21)	(.01)
Number of scenes	724.48***	4.41***	.01***
	(43.70)	(.41)	(.001)
Visual variation	-8,440.36	145.66	.23
	(9,283.33)	(86.93)	(.18)

Note: All tests are two-sided at * $p < .05$, ** $p < .01$, and *** $p < .001$.

Table 1. Voice Numerosity and Speech Rate in Crowdfunding (Study 1)

[1] The dependent variable in model 3, project success, is constructed based on the project's funding goal. Hence, funding goal is not included as an explanatory variable in this model.

3. Controlled Experiment: Voice Numerosity Under Varied Processing Resources

This study aims to provide more evidence that the effect is driven by enhanced cognitive elaboration. Participants whose processing capacity was varied were shown a product video and asked to indicate their maximum willingness to pay (WTP) for the product. Half of the participants watched a video with the same voice describing the product; the other half watched the same video but with five voices sequentially describing the product. We predicted that participants' WTP would be higher when the product information is narrated by more voices than when it is described by one voice, and this effect would be more pronounced when participants have more processing capacity (low distraction) than when they have limited processing capacity (high distraction).

3.1 Method

Participants and Design. A total of 382 participants (52% women; average age = 34.4) were recruited on TurkPrime online panel for a small monetary compensation. They were randomly assigned to one of four conditions of a 2 (distraction: high vs. low) × 2 (number of voices: 1 vs. 5) between-subjects design.

Pretest. The purpose of the pretest was to test the effectiveness of the number-of-speakers manipulation. One hundred TurkPrime participants (44% women; average age = 35.3) were randomly assigned to one of two conditions (number of voices: 1 vs. 5) in a between-subjects design. They were asked to watch a short video about a wireless charger and identify the number of voices they heard in the video.

To vary the number of voices, we created different versions of the product video. The spoken and visual contents are identical across different versions, except the voice(s) conveying the spoken content. In the one-voice condition, the product was described by the same voice throughout the video. In the five-voice condition, the product was described by a total of five voices throughout the video. To create different voices, we used five voice-synthesis models. The resulting synthetic voices were counterbalanced to ensure that the general quality was comparable and to rule out the possibility of order effect across conditions. We then created modified videos for the experiment, by combining the audio tracks with synthetic voices to the visuals of the original video. Results of the pretest confirmed that participants could identify the different number of voices created by the voice-synthesis models, with a median of 1 in the 1-voice condition and 5 in the 5-voice condition (Kruskal-Wallis, $\chi^2(1) = 58.64$, $p < .0001$).

Procedure and Measures. The main study was administered as two supposedly unrelated studies. The "first" study was an established operationalization to vary distraction (Nowlis & Shiv, 2005). In the high-distraction (vs. low-distraction) condition, participants were asked to memorize a 10-digit (vs. 2-digit) number.

In the "second" study, participants were asked to imagine that they were looking for a wireless charger for their cell phone and came across a new product in an online marketplace (Kickstarter). All participants watched a short product video, which was the same video as in the pretest. We varied the number of voices conveying the spoken content, but held the visual and spoken contents of the video constant.

As the main dependent measure, participants were asked to indicate their WTP for the product in the video on a 9-point scale from $30 to $70 (in $5-increments). As a check for the number of voices, participants reported the number of different voices speaking in the video by selecting a number from 1 to 7.

Participants then reported the number they memorized and indicated how easy or difficult it was to remember the number on a 7-point scale ("very easy/very difficult"); it was the check for distraction. Finally, they reported basic background information (gender, age), and their general interest with product innovations on crowdfunding platforms (1 = not at all interested; 7 = very interested).

3.2 Results

Twenty-five participants were removed for being "not at all interested" in innovations on crowdfunding (scoring 1 on reported general interest), leaving 357 observations. Further checks showed that the manipulations—number of narrating voices and distraction—were both successful, yielding only the predicted main effects in the respective ANOVA models.

An ANOVA of participants' WTP exhibited a number-of-speakers × distraction interaction ($F(1, 353) = 4.90$, $p = .027$). Under high distraction, participants had comparable WTP ($M_{1\text{-voice}} = \$44.81$, $M_{5\text{-voice}} = \$43.40$; $F < 1$). Under low distraction, participants' WTP were higher when the product was described by more speaker voices ($M_{5\text{-voice}} = \$47.98$, $M_{1\text{-voice}} = \$44.18$; $F(1, 353) = 5.34$, $p = .021$), exhibiting the hypothesized effect.

4. General Discussion

Across the studies, we show the hypothesized effect of voice numerosity and its boundary conditions, which support our conceptualization that the persuasive effect underlying more narrator voices is due to enhanced cognitive processing. Moreover, we find that the

measured effect size of voice numerosity is managerially significant. For instance, in study 1, our results showed that for each additional voice in the project video, the average project sees an increase of (1) about $12,795 in pledged amount (a 39% increase), (2) 118 customers supporting the project (backers) (a 38% increase), and (3) 1.6% probability that the project is successfully funded (a 6.5% increase). In Study 1, we find that the hypothesized effect is consistently moderated by the rate at which the spoken ad message was delivered. To the extent that faster speech rate disrupts cognitive processing (Moore et al., 1986), the results are consistent with the notion that voice numerosity is driven by enhanced processing. The next study tested the hypothesized effect using a controlled experiment. Results showed that the effect was moderated by consumers' processing resources. Together, the findings across both studies indicated that voice numerosity is facilitated by cognitive processing of the spoken message. Consistent with our conceptualization, across the studies, we find that the persuasive influence of voice numerosity is more pronounced (a) when the message is said at a slower rate; and (b) when consumers have more processing resources to process the message.

Voiceover narration is becoming even more popular and important to consumers in the modern marketplace. However, research on the effect of narrator voice on consumer behavior remains fairly under-researched (Krishna & Schwarz, 2014; Dahl, 2010). We aim to contribute to the sensory marketing by identifying and showing that the number of voices communicating a spoken message can affect consumer behavior. Moreover, the Marketing Science Institute (2020) has identified in its top research priorities the need for approaches "…to capture and analyze unstructured data such as video, voice, and text in order to improve firm communications and customer experience" (p. 9). Our research speaks to this issue. We aim to contribute to practice by offering concrete recommendations on voiceover narration for marketing practitioners and architects of the consumer information environment (e.g., UX designers, entrepreneurs) to consider in designing persuasive communications. Our findings suggest that for product messages that are simpler to understand, it may be worthwhile to have multiple narrators; for more complex product messages, it might be more effective to have just one narrator. Moreover, practitioners can take into account the speech rate in delivering the narrated message, when leveraging the voice numerosity effect.

References.

Cacioppo, J. T., Petty, R. E., & Chuan Feng Kao. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment*, *48*(3), 306. https://doi.org/10.1207/s15327752jpa4803_13

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*, 975–979. https://doi.org/10.1121/1.1907229

Dahl, D. W. (2010). Understanding the role of spokesperson voice in broadcast advertising. In A. Krishna (Ed.), *Sensory marketing: Research on the sensuality of products* (pp. 169–182). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780203892060-21

Dhanani, Q., & Mukherjee, A. (2017). *Is crowdfunding The silver bullet to expanding innovation in the developing world?* World Bank Blogs. Retrieved December 1, 2021, from https://blogs.worldbank.org/psd/crowdfunding-silver-bullet-expanding-innovation-developing-world.

Fan, T., Gao, L., & Steinhart, Y. (2020). The Small Predicts Large Effect in Crowdfunding. *Journal of Consumer Research, 47*, 544-565. https://doi.org/10.1093/jcr/ucaa013

Goldberg, M.E., Chattopadhyay, A., Gorn, G.J. and Rosenblatt, J. (1993), Music, music videos, and wear out. *Psychology & Marketing,* 1, 1-13. https://doi.org/10.1002/mar.4220100102

Krishna, A., & Schwarz, N. (2014). Sensory marketing, embodiment, and grounded cognition: A review and introduction. *Journal of Consumer Psychology*, *24*(2), 159–168. https://doi.org/10.1016/j.jcps.2013.12.006

Lynch Jr., J. G., & Srull, T. K. (1982). Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods. *Journal of Consumer Research*, 9(1), 18–37. https://doi.org/10.1086/208893

Li, Xi, Mengze Shi, and Xin Wang (2019), "Video Mining: Measuring Visual Information Using Automatic Methods," *International Journal of Research in Marketing,* 36 (2), 216–31.

Liu, Xuan, Savannah Wei Shi, Thales Teixeira, and Michel Wedel (2018), "Video Content Marketing: The Making of Clips," *Journal of Marketing,* 82 (4), 86–101.

Makino, T., Liao, H., Assael, Y., Shillingford, B., Garcia, B., Braga, O., & Siohan, O. (2019). Recurrent neural network transducer for audio-visual speech recognition. In *2019 IEEE automatic speech recognition and understanding workshop (ASRU)*, 905-912. https://doi.org/10.1109/ASRU46091.2019.9004036

Marketing Science Institute. (2020). *MSI Announces 2020-22 Research Priorities*. Retrieved from https://www.msi.org/articles/2020-22-msi-research-priorities-outline-marketers-top-concerns/ (Last accessed: December 1, 2021)

Moore, D. L., Hausknecht, D., & Thamodaran, K. (1986). Time Compression, Response Opportunity and Persuasion. *Journal of Consumer Research*, *13*(1), 85–99. https://doi.org/10.1086/209049

Nowlis, S. M., & Shiv, B. (2005). The Influence of Consumer Distractions on the Effectiveness of Food-Sampling Programs. *Journal of Marketing Research (JMR), 42*(2), 157–168. https://doi.org/10.1509/jmkr.42.2.157.62287

Pennebaker, J.W., Boyd, R.L., Jordan, K.N., & Blackburn, K.G. (2015). The Development and Psychometric Properties of LIWC2015.https://doi.org/10.15781/T29G6Z

Younkin P., & Kuppuswamy, V. (2018). The Colorblind Crowd? Founder Race and Performance in Crowdfunding. *Management Science, 64*(7), 3269–3287. https://doi.org/10.1287/mnsc.2017.2774