How the Voice Persuades

Alex Van Zant, Rutgers University, USA
Jonah Berger, University of Pennsylvania, USA

While persuasion attempts executed through linguistic channels often backfire, little is known about paralinguistic persuasion, or persuasion attempts executed through nonverbal properties of communicators’ voice. Across three experiments, we find that paralinguistic attempts enhance persuasion because they signal confidence. Speakers’ greater volume and variance in volume shaped the persuasion process.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/2411691/volumes/v46/NA-46

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Finally, Kumar and Epley examine how the voice shapes social connection. When interacting for the first time, or reconnecting with old friends or people tend to prefer text because they think voice-based interactions will be more awkward. They’re incorrect. There are actually no differences in awkwardness between communication media, but anticipated awkwardness drives people to use text-based media to interact, decreasing social connection.

Taken together, these papers provide convergent evidence for the value of voice. Technology has changed how people communicate. More and more people now used text-based media (i.e., text, email, or social media) rather than talking on the phone or face-to-face. But as these four papers show, there are a number of important benefits to voice (e.g., confidence, sincerity, humanization, and connection) that help voice increase persuasion, attitude change, and social bonding.

This session should be of interest to a range of consumer behavior researchers, from those that study social influence and persuasion to those interested in prosocial behavior, social interaction, political discourse, and well-being.

How the Voice Persuades

EXTENDED ABSTRACT

Persuading others can be challenging. When people diagnose others’ intent to persuade, they often defend themselves by ignoring or counter arguing against the message (Friestad & Wright, 1994). Consumers can diagnose common persuasion tactics like positively valenced statements (e.g., “exceeds your needs” versus “meets your needs”), for example, or rhetorical questions (e.g., “Mizuno shoes are beneficial for you, aren’t they?”), which causes communicators’ persuasion efforts to backfire (Campbell & Kirmani, 2000).

When studying persuasion, however, the literature have often examined written messages. It has looked at manipulating written arguments (Tormala & Petty, 2002), print advertisements (Kirmani & Zhu, 2007), and text-based hypothetical vignettes (Campbell & Kirmani, 2000).

In contrast, little work has examined persuasion attempts executed through paralinguistic channels, or nonverbal properties of communicators’ voice. Communicators can deliver the same message with in different ways, varying acoustic properties like their pitch, volume, and speech rate. Might such shifts actually make communicators more persuasive?

We argue that, although communicators may struggle to persuade others through what they say (linguistic persuasion attempts), they may be relatively effective at persuading others through how they say it (paralinguistic persuasion attempts).

If paralinguistic persuasion attempts are effective, there are two possible reasons why. The first is that listeners fail to detect them, which may make them harder to resist. Whereas linguistic attempts are fairly easy for listeners to identify (and thus ignore or react against), paralinguistic persuasion cues may be more difficult to diagnose. Alternatively, paralinguistic persuasion attempts may increase persuasion through signaling confidence. Confidence is a credibility cue that enhances persuasion (Karmarkar & Tormala, 2009). If speakers sound more confident when they are trying to persuade, this may boost persuasion regardless of whether their attempts are detected.
Across three experiments, we examine (1) whether paralinguistic persuasion attempts enhance persuasion and (2) the mechanisms underlying this effect.

Experiment 1-3. First, we generated stimulus materials by recording a sample of speakers ($N = 49$) reading a review of a TV show twice (randomized order): once when incentivized to persuade a future research participant (paralinguistic attempt condition) and once when not given such an incentive (control condition).

Second, using a (paralanguage: control, persuasion) × 2 (disclosure statement: no, yes) between-subjects design, we presented a separate sample ($N = 713$) with a randomly selected recording from the stimulus materials. To test whether paralinguistic persuasion attempts work because they are less likely to be detected, we also manipulated whether listeners received information about speakers’ intent to persuade (i.e., the presence or absence of a disclosed monetary incentive to persuade).

Finally, we measured whether participants were persuaded (4 items; e.g., “I would like to purchase this TV”) and perceived the speaker as confident (2 items: confident, certain).

As predicted, paralinguistic attempts enhanced persuasion ($p = .049$). This was not moderated by linguistic information about communicators’ intent to persuade, however, indicating that even when persuasion was detected, paralinguistic attempts still enhanced persuasion. Instead, results are more consistent with a confidence account. Paralinguistic attempts made speakers appear more confident ($p = .003$), which mediated the effect of paralinguistic attempts on persuasion.

While one could argue that the results of Experiment 1 are somehow driven by communicators’ intent to persuade not being clear enough, Experiment 2 ($N = 1,104$) finds the same effects on persuasion ($p = .003$) and confidence ($p < .001$), including mediation by confidence, even when communicators directly acknowledge their intent to persuade (i.e., “I would like to convince you to purchase this TV”).

In Experiment 3, we allowed speakers ($N = 45$) to craft both the verbal and nonverbal aspects of their appeals. We manipulated linguistic persuasion attempts by randomly assigning half just to compose a review (control condition) and asking the other half to compose a persuasive review (linguistic attempt condition). Then, as in the first two experiments, they read their review aloud twice, once while making a paralinguistic persuasion attempt.

Next, we presented a randomly selected recording from the stimulus materials to a separate sample of research participants ($N = 1,086$) and assessed persuasion and confidence. Finally, to provide a stronger test of detectability, we directly assessed the extent to which participants could diagnose communicators’ intent to persuade (“I thought it was pretty obvious that the speaker was trying to persuade me”).

As predicted, paralinguistic attempts again enhanced persuasion ($p < .001$). This occurred even through listeners accurately perceived speakers’ greater persuasive intent ($p < .001$), providing further evidence against detectability. Instead, the results again support the confidence account. Paralinguistic attempts made speakers appear more confident ($p < .001$), which mediated the effect of paralinguistic attempts on persuasion.

In contrast, linguistic persuasion attempts slightly harmed persuasion ($ns$) and did not make speakers appear more confident despite being similarly detectable to paralinguistic attempts ($p = .037$). This suggests that although participants could detect them, paralinguistic attempts were more effective than linguistic attempts because they signaled confidence.

**Exploratory Analysis of Speakers’ Paralinguistic Cues.** To better understand how speakers’ paralinguistic attempts enhanced persuasion, we also analyzed the recordings used in each experiment to explore which nonverbal vocal cues influenced the persuasion process. We extracted objective phonetic measures using the Praat program (Boersma & Weenik, 2017). Relative to the control condition, speakers increased their volume ($p < .001$) and varied their volume to a greater extent ($p = .014$) when attempting to persuade. Each of these cues enhanced their perceived confidence and, in turn, the extent to which they were persuasive.

**Conclusion.** Taken together, these findings suggest that because they signal confidence, paralinguistic persuasion attempts are effective even when they are diagnosed. Beyond *what* to say, focusing on *how* to say it can increase influence.

**Sincere Persuasion:**

Cues in Language Used to Assess and Convey Sincerity

**EXTENDED ABSTRACT**

Sincerity is a valued aspect of personality (Anderson, 1968), emotional displays (Grandey, Fisk, Mattila, Jansen, & Sideman, 2005), and communication (Barasch, Berman, & Small, 2016). Sincerity can be persuasive because it assuages any concern about ulterior motive (Campbell & Kirmani, 2000). Furthermore, sincerity is a key input into many consumer decisions, such as whether to take a recommender’s advice about purchasing a product (Tuk, Verlegh, Smidts, & Wigboldus, 2009).

But what determines whether persuaders are deemed sincere? The current research examines the language cues that people use to judge others’ sincerity.

In particular, we focus on three primary sets of language cues that could be used to evaluate sincerity: semantic cues (i.e., the message content or actual words expressed), auditory cues (i.e., characteristics of the communicator’s voice, such as tone or volume), and visual cues (i.e., characteristics of the communicator’s visual appearance, such as facial expression). While previous research suggests that each of these cues affects judgments of a message (semantic: Packard & Berger, 2017; auditory: Schroeder, Kardas, & Epley, 2017; visual: De Waale & Claes, 2017), it is unclear how important each cue is for assessing sincerity specifically, and how these cues interact to influence perceptions of a communicator. Our work tests this question in the context of persuasive charity appeals.

First, to examine consumers’ lay beliefs about which cues best convey sincerity, we conducted an online survey ($n = 104$) asking participants how effective different communication cues are for conveying sincerity (i.e., to support a cause). They evaluated three sets of cues: hearing a communicator’s appeal (auditory cues), reading a communicator’s appeal (semantic cues), or watching a communicator’s appeal (visual cues). Participants believed auditory cues would be the most effective ($ps < .002$), and that semantic and visual cues would be equally less effective ($p = .757$). The same pattern of results emerged for ratings of how useful each medium would be for determining sincerity in others and expressing sincerity to others.

Next, we test whether lay people are correct by conducting three experiments. In each experiment, we used real charity appeals (videotaped; $n = 36$ in Study 1; $n = 93$ in Study 2; $n = 76$ in Study 3; see Barasch et al., 2017 for a prior analysis of donations using these video appeals). We then manipulated the communication medium that observers evaluated. In Study 1 ($n = 678$), we randomly assigned observers to 1) watch only the silent video (visual condition), 2) listen to only the audio (audio condition), 3) read only the content of the appeal (text condition), or 4) watch the video and audio (full
condition). Each observer evaluated only one condition from one speaker, and subsequently assessed the speaker’s sincerity on six pre-validated items (a = .94). Participants’ mean-level assessments of speakers’ sincerity were highest in the full and audio conditions compared to the other conditions. Additionally, a follow-up analysis revealed that observers’ sincerity ratings in the full condition were predicted most strongly by observers’ ratings in the audio condition (β = .63, p < .001), followed by the text condition (β = .53, p < .001), and least strongly in the visual condition (β = .16, p = .35).

The results from Study 1 show that hearing someone’s voice signals sincerity significantly above reading their words or seeing them deliver an appeal. However, the comparison between auditory and visual cues in Study 1 contains a confound: a person’s voice contains both semantic and auditory cues whereas a silent video contains only visual cues without semantic information. To determine whether people also attend to visual information when semantic content is present, Studies 2 (n = 1,376) and 3 (n = 2,262, pre-registered at http://aspredicted.org/blind.php?x=fr88n2) added a fifth experimental condition: watching video with subtitles (“subtitled video” condition). Results from both studies showed that subtitled videos added significantly more predictive validity above silent videos alone (Study 2: p = .006; Study 3: p < .001) and text alone (Study 2: p = .01; Study 3: p = .001). These studies also replicated Study 1’s finding that audio ratings predict full ratings above and beyond text ratings (Study 2: p = .04; Study 3: p = .02).

To better understand the pattern of results across all three studies, we conducted an analysis with all participants (n = 4,316) controlling for study. This analysis revealed three key findings. First, observers attributed the greatest sincerity to persuaders in the audio condition compared to persuaders in the text and subtitled video conditions, and the least sincerity to the silent video condition. That is, hearing speakers—but not seeing speakers—meaningfully increased observers’ evaluations of people’s sincerity. This might suggest that auditory cues convey greater sincerity—or less insincerity—than visual cues, or that people naturally tend to use their voice to express sincerity more than their face or body.

Second, observers relied most strongly on semantic content and auditory cues to assess sincerity; they relied relatively less strongly on visual cues. This suggests that a sincere voice may be more powerful for conveying sincerity than a sincere face.

Third, semantic content combined with non-semantic cues, either auditory or visual, predicted sincerity judgments better than semantic content alone. One implication of this finding is that the words people use are significantly limited in conveying sincerity in the absence of vocal or visual cues. Rather, it is the combination of semantic and paralinguistic cues that make one’s words seem sincere.

In conclusion, four experiments demonstrate that auditory cues may be particularly important for judging others’ sincerity, and visual cues relatively less important. While lay people recognize the significance of auditory cues, they overestimate the significance of visual cues. This research has implications for many marketing contexts where successful persuasion depends on perceptions of sincerity, from salespeople convincing customers that clothing looks good on them, to celebrities convincing the public they are sorry for a transgression, to doctors convincing patients to take certain medications. If sincere, persuaders should prioritize being heard over being seen.

Expressing Dissent: How Communication Medium Shapes Dehumanization and Attitude Change

EXTENDED ABSTRACT

Disagreement often begets incivility. When one person disagrees with another, they tend to infer not only that the other person has deficient opinions but also deficient mental capacities (Kennedy & Pronin, 2008; Pronin, Lin, & Ross, 2002). In such a way, people tend to dehumanize those with whom they disagree: believing that their intellectual faculties are weaker, with mental capacity more like a thoughtless animal than a thoughtful human (Epley & Waytz, 2010). Some advocates for civil discourse, therefore, have suggested trying to “humanize” interactions between ideological opponents, for example by having opponents exchange opinions in person instead of online. Can changing the communication medium by which dissenting opinions are exchanged affect dehumanization, thereby influencing attitude polarization?

Prior research suggests at least two ways in which the interaction medium could affect dehumanization. First, the medium can affect a person’s message—that is, the words he or she selects to use in the conversation. Communicators may express more antagonistic and aggressive opinions online, because online interactions increase disinhibition in part because they create emotional distance (e.g., Suler, 2004). However, writing may provide communicators with more time to carefully compose their thoughts than speaking; communicators may write more eloquently than they speak (Hayes, 1988).

Second, beyond affecting how a message is generated, the medium affects how a message is consumed. For example, hearing a person’s opinions compared to reading the same opinions makes him or her seem more mentally sophisticated and is hence humanizing, in part because cues in a person’s voice such as tone and pauses convey their intellect (Schroeder & Epley, 2015, 2016; Schroeder, Kardas, & Epley, 2017).

To test how communication medium affects humanization and attitude change, Experiment 1 recruited 396 political opponents to exchange opinions for eight minutes in a pre-registered experiment. Participants first completed a pre-survey measuring their opinions on three pre-tested contentious political and social topics (e.g., the use of GMOs), were then paired based on disagreement, and finally randomly assigned to either have a spoken or written conversation. After meeting their partner, speaking-condition pairs went to a private room to have their conversation (which we recorded) whereas writing-condition pairs went to a computer lab and had their conversation online.

When the conversations were complete, participants completed a post-survey evaluating their partner (measuring humanization using a pre-validated scale; Bastian & Haslam, 2010) and the conversation and reporting their updated attitudes. Results revealed an effect of communication medium on humanization: opponents humanized each other more when speaking (M=5.21 on 7-point scale) than when writing (M=5.67), F=16.38, p<.001, η²=.078. Moreover, they reported feeling more understood by their partner, enjoyed the conversation more, and reported more attitude change, F=24.82, p<.001, η²=.113. The effect of speaking vs. writing on attitude change was mediated by humanization. But they experienced no different levels of conflict during the conversation, p=.250 (i.e., ratings of how emotional, heated, contentious, and conflicted was the conversation).

The effect of interaction medium on humanization could be driven either by different types of messages being created in each medium or by differences in how the messages are consumed. In Experiment 1, the “creation medium” and “consumption medium”
were conflated, making it impossible to disentangle which aspects of
the medium affected humanization. To test between the effect of
the creation and consumption medium, we created transcriptions of
the spoken conversations and asked research assistants matched on
gender to read aloud the written conversations (recorded). In Experi-
ment 2, online observers (n=994) either listened or read one of the
conversations and evaluated the mental capacities of each interactant
on the same humanization scale that the original participants used.
Observers also reported their own opinions on the topics. Therefore,
Experiment 2 used a 2 (creation medium: speaking vs. writing) × 2
(consumption medium: hearing vs. reading) design: Observers either
heard or read an originally spoken or written conversation.

Two main effects emerged (no interaction). First, regardless of
how the conversation was created, observers humanized communicators
more when they heard the conversation than when they read it,
F=4.41, p=.036, η²=.004, suggesting that a person’s intellect is more
apparent when their voice is heard than when their words are read.

But second, the effect of the creation medium indicated that
written statements humanized the communicator more than spoken
statements: Observers rated the communicators who wrote their
opinions as being more intelligent and consequently more uniquely
human than the communicators who spoke their opinions, F=10.25,
p<.001, η²=.03. This suggests that, at least in this context, communica-
tors were better able to convey their intellect by writing than by
speaking.

To further test whether communicators exchanged less an-
tagonistic or contentious messages when speaking than writing, we
restricted our observer sample to only those who disagreed with
communicators’ opinions. Again, observers humanized the
communicators who wrote (vs. spoke) marginally more highly, F=2.87,
p=.090, η²=.005, suggesting that communicators did not create
more antagonistic messages when typing.

Overall, these results suggest that how a message is consumed
may matter more for evaluations of an opponent than how it was
created.

In summary, how people interact influences humanization,
which in turn shapes attitude change. Opponents humanized each
other more during in-person (spoken) interactions than online (wri-
ten) interactions and consequently reported more change in attitudes.
There was no empirical evidence that individuals create more intel-
ligent or less divisive messages when they speak (vs. write) to each
other. Instead, consistent with prior research (e.g., Schroeder et al.,
2017), the cues in a person’s voice may convey their intellect more
clearly than do their words alone. In other words, how a person’s
opinions are consumed (heard or read) may be even more important
for humanization than how the opinions are generated (via speaking
or writing).

These findings have important implications for how technol-
yogy may shape discourse. People who disagree can easily interact
via online forums (e.g., Reddit, Facebook), which may discourage
them from conversing in-person. This work suggest that these online
interactions may subtly enhance dehumanization. Disagreement is
born not just from harsh words exchanged but from the structure of
the interaction itself.

**It’s Surprisingly Nice to Hear You: Misunderstanding the
Value of Talking (vs. Typing)**

**EXTENDED ABSTRACT**

Social connections improve consumer welfare (e.g., Kahneman
& Deaton, 2010). Technological advances offer many different ways
of connecting with others across different communication media.

But do consumers choose the medium through which they connect
wisely?

We predicted that people would feel a stronger bond with an-
other person when they talked than when they typed, because inter-
actions involving the human voice contain cues not present in text-
based media (Kruger, et al., 2005; Schroeder & Epley, 2015), cues
that may affect connectedness and that must be anticipated if con-
sumers are to choose media for engagement optimally. We further
predicted that people may not recognize the value of voice-based
media for promoting connection, and that preferences for communi-
cation media would instead be guided by the expected awkwardness
of an interaction.

To test these predictions, we ran an initial study with 300 par-
ticipants who completed questionnaires on the Friday before New
Year’s Eve and the following Monday. At both points, participants
rated how lonely they felt (Russell, 1996). In the follow-up survey,
participants reported how much time they spent interacting with oth-
ers in voice-based exchanges (face-to-face, phone, video chat) and
in text-based exchanges (email, texting, social media). The more
time spent in voice-based interactions, the less lonely participants
felt after the weekend compared to before, p < .01. The amount of
time spent on text-based communication was not related to changes
in loneliness.

Experiment 1 asked participants to reconnect with an old friend
they were once close to, but with whom they had fallen out of touch.
Participants predicted how connected and awkward they would feel
if they reconnected in two ways: over the telephone (using their
voice) and over email (using text). They also indicated which option
they preferred. Rather than give participants their chosen medium,
we randomly assigned them to either reconnect over the phone or
e-mail.

Participants expected they would feel like they had formed a
stronger bond with their friend over the phone than over email, and
also anticipated that they would feel like they had “really connected”
more strongly over the phone, p < .0001. However, participants also
expected that they would feel more awkward over the phone, p < .0001.
Sixty-seven percent of participants indicated that they would
prefer to interact over email than over the phone, χ²(N=200) = 23.12,
p < .0001. A regression predicting choice of communication media
from expected connection and awkwardness revealed the strongest
effect of expected awkwardness (p < .0001).

When they actually reconnected, however, participants report-
ed feeling a stronger bond when assigned to reconnect with an old
friend over the phone than over email, p < .05, and also reported a
stronger feeling that they “really connected” over the phone, p = .06.
In contrast, participants did not feel any more awkward when they
connected over the phone, p > .7.

Although the results of this naturalistic field experiment suggest
that the presence of voice could be producing these effects, to more
cleanly test our predictions, we designed a laboratory experiment
that ensured perfect response rates, constrained topics of discussion,
and more carefully controlled the media through which participants
interacted. In Experiment 2 (N=302), participants engaged in a modi-
fied “fast friends” procedure (Aron, et al., 1997) and chatted with a
stranger on Skype, either over video, voice, or text. The video condi-
tion allowed us to test whether providing visual cues would increase
connection compared to voice alone, or whether the voice uniquely
provides a sense of social connection. Before interacting, partici-
pants predicted how connected they would feel to their partner and
how awkward they expected the interaction to be. Unlike Experi-
ment 1, in which participants predicted their experience in both the
voice-based and text-based conditions, predictions here were made
between-subjects. Participants in this experiment predicted no differences across communication media in either connection or awkwardness (all ps > .1).

Participants, however, again felt more connected when they communicated in a voice-based medium than when they only used text. Adding additional cues beyond voice in the video condition did not increase connection compared to the audio condition, suggesting that it is not simply the addition of more interpersonal cues that increases the sense of connection between consumers. We also observed no differences in experienced awkwardness between communication media conditions (ps > .2). Because we collected predicted and actual ratings from participants for only the between-subjects condition to which they were assigned, we can directly compare predicted and actual experiences. Participants across all conditions significantly overestimated how awkward their interaction would be, p < .001, and underestimated how connected they would feel, p < .001. Moreover, the difference between predicted and actual ratings of connection was greater in the voice and video conditions than in the text condition, ps < .05. Note, though, that the gap between expected and experienced connection did not differ between the voice and video conditions, p > .8. These results suggest that connecting over voice-based media creates an especially surprising bond with another person.

A final experiment (N=50) examined whether predictions of connection or awkwardness are more likely to guide decisions in daily life by having consumers consider a variety of scenarios. For each situation, participants indicated how connected they would feel and how awkward it would be to interact over the phone and over email. They then indicated the extent to which they would rather communicate by talking or typing. Preferences were better predicted by expectations of awkwardness than connectedness. Recall, however, that our prior experiments revealed no differences in actual awkwardness across communication media.

How a consumer decides to interact with another person has important consequences for how he or she feels towards that person. These results suggest that miscalibrated expectations of awkwardness could lead people to choose communication media that may not optimize their own wellbeing. This research sheds light on how consumers choose to interact with each other and the consequences of their choices. Understanding these consequences is necessary in order to choose effectively. Voice-based media increase connection without the expected costs of awkwardness.

REFERENCES


