PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Superficial auditory (dis)fluency biases higher-level social judgment

Robert Walter-Terrill[a,1], Joan Danielle K. Ongchoco[a,b], and Brian J. Scholl[a,c,1]

Edited by Elke Weber, Princeton University, Princeton, NJ; received July 29, 2024; accepted January 2, 2025

When talking to other people, we naturally form impressions based not only on what they say but also on how they say it—e.g., how confident they sound. In modern life, however, the sounds of voices are often determined not only by intrinsic qualities (such as vocal anatomy) but also by extrinsic properties (such as videoconferencing microphone quality). Here, we show that such superficial auditory properties can have surprisingly deep consequences for higher-level social judgments. Listeners heard short narrated passages (e.g., from job application essays) and then made various judgments about the speakers. Critically, the recordings were modified to simulate different microphone qualities, while carefully equating listeners' comprehension of the words. Though the manipulations carried no implications about the speakers themselves, common disfluent auditory signals (as in "tinny" speech) led to decreased judgments of intelligence, hireability, credibility, and romantic desirability. These effects were robust across speaker gender and accent, and they occurred for both human and clearly artificial (computer-synthesized) speech. Thus, just as judgments from written text are influenced by factors such as font fluency, judgments from speech are not only based on its content but also biased by the superficial vehicle through which it is delivered. Such effects may become more relevant as daily communication via videoconferencing becomes increasingly widespread.

decision-making | perceptual fluency | social judgments

Significance

In recent years, tools such as videoconferencing have shifted many conversations online, with stark auditory ramifications—such that some voices sound clear and resonant while others sound hollow or tinny, based on microphone quality and characteristics. A series of experiments shows that such differences, while clearly not reflective of the speakers themselves, nevertheless have broad and powerful consequences for social evaluation, leading listeners to make lower judgments of speakers' intelligence, hireability, credibility, and even romantic desirability. Such effects may be potential sources of unintentional bias and discrimination, given the likelihood that microphone quality is correlated with socioeconomic status. So, before joining your next videoconference, you may want to consider how much a cheap microphone may really be costing you.

Our daily interactions with others are naturally influenced by our assessments of their underlying traits (e.g., trustworthiness, intelligence) (1), and since we do not typically have access to their full life histories, we must (and do) make such judgments on the basis of surprisingly small snapshots of their behavior—so-called "thin slices" (2–4). For example, even within the initial 30 s of a new conversation, people form numerous impressions about the speakers (5)—and such judgments are based not only on what they say (6) but also on how they say it in the first place (7). Intuitively, the sounds of people's voices can reveal states or traits, such as trustworthiness (8, 9), which can be formed within as little as 400 ms (10). And some intuitions may sometimes even be valid: For example, since vocal tract length is tightly correlated with body size, both in humans (11) and across the animal kingdom (12), we can effectively hear how large (11) or how strong (13) an organism is by simply listening to it. (It is physically impossible for a chihuahua to produce a lion's growl, and vice versa.) Similarly, the pitch of a speaker's voice can serve as a valid cue to their personality (14) and can even influence their likelihood of winning elections (15).

In the current era, however, how one sounds to others is often determined not only by relevant intrinsic properties (such as vocal anatomy) but also by extrinsic and superficial properties that may be entirely irrelevant to subsequent character assessments—such as the quality of one's computer microphone. In fact, our interest in auditory fluency was motivated by our real-world experiences with different types of auditory signals due to differing microphone qualities, in the context of videoconferencing. Some participants' voices sound so clear, present, and lively that they seem to be speaking from a recording studio, while others sound hollow and tinny, as might happen when speaking through the microphone of a cheap older laptop. Critically, these differences seem salient even when the actual words are equally comprehensible in both cases. This particular context of videoconferences may once have seemed somehow specialized or boutique, but of course, it has become much more relevant and widespread since the COVID-19 pandemic forced much of education, business, commerce, government, and even daily conversation to happen online. Here, we demonstrate that such differences in auditory quality have surprisingly deep implications for higher-level social judgments.

Visual and Auditory Fluency

This study was inspired by work on fluency (the ease, or lack thereof, of information processing) in the visual domain. Suppose you read a written personal statement and then must judge the author's intelligence. Of course, you might naturally base your judgment

Author affiliations: [a]Department of Psychology, Yale University, New Haven, CT 06520-8047; [b]Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; and [c]Wu-Tsai Institute, Yale University, New Haven, CT 06520-8047

Author contributions: R.W.-T., J.D.K.O., and B.J.S. designed research; R.W.-T. performed research; R.W.-T. and J.D.K.O. analyzed data; and R.W.-T., J.D.K.O., and B.J.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: robert.walter@yale.edu or brian.scholl@yale.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2415254122/-/DCSupplemental.

Published March 24, 2025.

on multiple relevant intrinsic factors—e.g., the quality of the arguments, or the sophistication of the vocabulary. But such judgments are also influenced by extrinsic factors shaped by the message's "vehicle" rather than its content, such as the font in which the passage is written. Indeed, "disfluent" fonts—those that are perfectly legible but require more time and effort to read—lead to lower intelligence ratings than do more fluent fonts that are easier to read (16). Such effects can be both powerful and ubiquitous—influencing everything from predicted course difficulty based on syllabus fonts (17) to predicted food quality based on menu fonts (18). More generally, fluency has robust effects on our mental lives, from perception, metacognition, and language, to decision-making, imagery, and memory—while shaping judgments relating to truth, confidence, frequency, preference, valuation, causation, and character traits (19–22), and these cues are often valid and informative (23, 24). Fluency effects are thought to be driven by a series of affective and cognitive mechanisms, beginning with a) a subjective experience from the ease (or lack thereof) of processing (22) which b) serves as a metacognitive cue, leading to an attribution (or misattribution) of that cue to the target being evaluated, and c) ending with a contextualized interpretation of the cue based on learned associations and naive theories (25).

As with most perceptual phenomena, studies of fluency have tended to focus on visual manipulations—with work in the auditory domain often involving audiovisual stimuli (26), or focusing not on extrinsic properties of speech (such as the impact of microphone quality), but rather on intrinsic properties that may be directly relevant to the judgments at hand. For example, teachers who speak haltingly and read from notes are judged to be less effective by students in a classroom context (27). Further, artificial pauses introduced into speech recordings are readily misattributed to the speakers themselves and can influence judgments of states such as attentiveness (28). In two innovative studies that focused most directly on extrinsic auditory manipulations, disfluent audio in the context of science communication was judged less favorably (29), and disfluent witness testimony in a mock legal proceeding was found to be less trustworthy and led to poorer memory of key facts (30). However, the disfluent audio in these studies was distorted in extreme and unfamiliar ways that compromised the comprehensibility of the speech itself [e.g., with only 64% of subjects "correctly respond(ing) about the content of … talks" (29, p. 255), or nearly 20% of subjects being excluded for not being able to explain what the testimony was about (30)]. In contrast, the current study explored highly familiar forms of auditory (dis)fluency, while equating word-for-word comprehensibility. In this sense, we aim in this study to isolate effects of auditory fluency to the superficial vehicle of the speech, without influencing the perception of the speech's content—and we aim to do so in contexts that most readers will experience frequently in their daily lives.

The Current Study

We explored whether superficial and extrinsic properties of the speech signal itself—independent of the actual message—would impact higher-level social judgments. Listeners heard a speech recording and then made a single judgment about the speaker. For example, in some experiments, listeners heard a putative job applicant reading a short passage from a personal statement, after which the listeners judged "the likelihood that you would hire this person". Critically, half of the listeners heard a clear recording (made from a high-quality microphone), while the other half heard a distorted version (mimicking familiar poor microphone quality), utilizing a between-subjects design. (Much past work in this domain has employed within-subjects fluency manipulations, but we wanted to explore such effects without introducing task demands that could call attention to them—and to foreshadow, this was successful, e.g., with only 0.67% of our subjects in one key experiment suspecting that the study had anything to do with auditory signal quality.) The distortions resulted in hollow or "tinny" voices that will be immediately familiar to anyone with videoconferencing experience but are clearly extrinsic insofar as they are unreflective of what voices sound like during in-person conversation (Audios S1–S8). Critically, while some past work on auditory quality has intentionally rendered speech unintelligible (7), the manipulations in the current experiments did not impact the brute comprehensibility of the words and sentences themselves. [In contrast, properties such as perceived accents (31–34) can influence judgments by reducing comprehension and might reflect intergroup stereotypes and biases, beyond fluency effects (35)].

Across six preregistered experiments, we explored such effects while varying both speaker gender (male vs. female) and accent (American vs. British English). Additionally, we tested conditions not only with human voices but also with clearly computer-synthesized voices, which cannot reflect any intrinsic qualities of the speaker. (And indeed, since we applied the very same manipulations to the spoken experimental instructions themselves, the speakers clearly could not be responsible for them.) We explored how such effects may impact several different types of higher-level social judgments—including both abstract trait-like measures (of credibility and intelligence) and more concrete real-world decisions (of hireability and romantic desirability). Finally, we examined the scope of such effects to determine whether they are specific to judgments of the speakers or if they might instead reflect more generalized changes in judgment and affect.

Results

Hireability. We first tested for the effect of disfluent auditory signals in the context of an especially concrete decision (Experiment 1). Listeners were asked to imagine that they were tasked with making a hiring decision for a highly competitive "senior sales manager" position and then heard a purported personal statement from a potential applicant, recorded in a human male voice (Audios S1 and S2). When asked to rate how likely they would be to hire this person on a continuous scale from "Very Unlikely" to "Very Likely", listeners who heard the Distorted recording (mimicking a poor-quality microphone) indicated (as depicted in Fig. 1*A*) that they were less likely to hire the speaker compared to those who heard the Clear recording: 68.50 (SD = 24.07) vs. 76.04 (SD = 18.98), $t(598) = 4.26$, 95% CI [4.07, 11.02], $P < 0.001$, d = 0.35. This effect occurred despite the fact that subsequent word-for-word transcription accuracy did not differ across the two recordings: 98.07% (SD = 0.03) vs. 98.09% (SD = 0.03), $t(598) = 0.11$, 95% CI [0.00, 0.00], $P = 0.915$, d = 0.01, $BF_{10} = 0.091$; coder agreement 97.79%.

Desirability. Beyond hireability in a professional context, we also asked whether such auditory fluency effects would generalize to even more socially contextualized judgments, of romantic desirability (Experiment 2). An independent group of listeners was asked to imagine that they were single and were looking for a potential date on an online dating site and then heard a statement of interest, recorded in a human female voice (Audios S3 and S4). When asked to rate how likely they would be to go on a date with this person on a continuous scale from Very Unlikely to Very Likely, listeners who heard the Distorted recording indicated (as

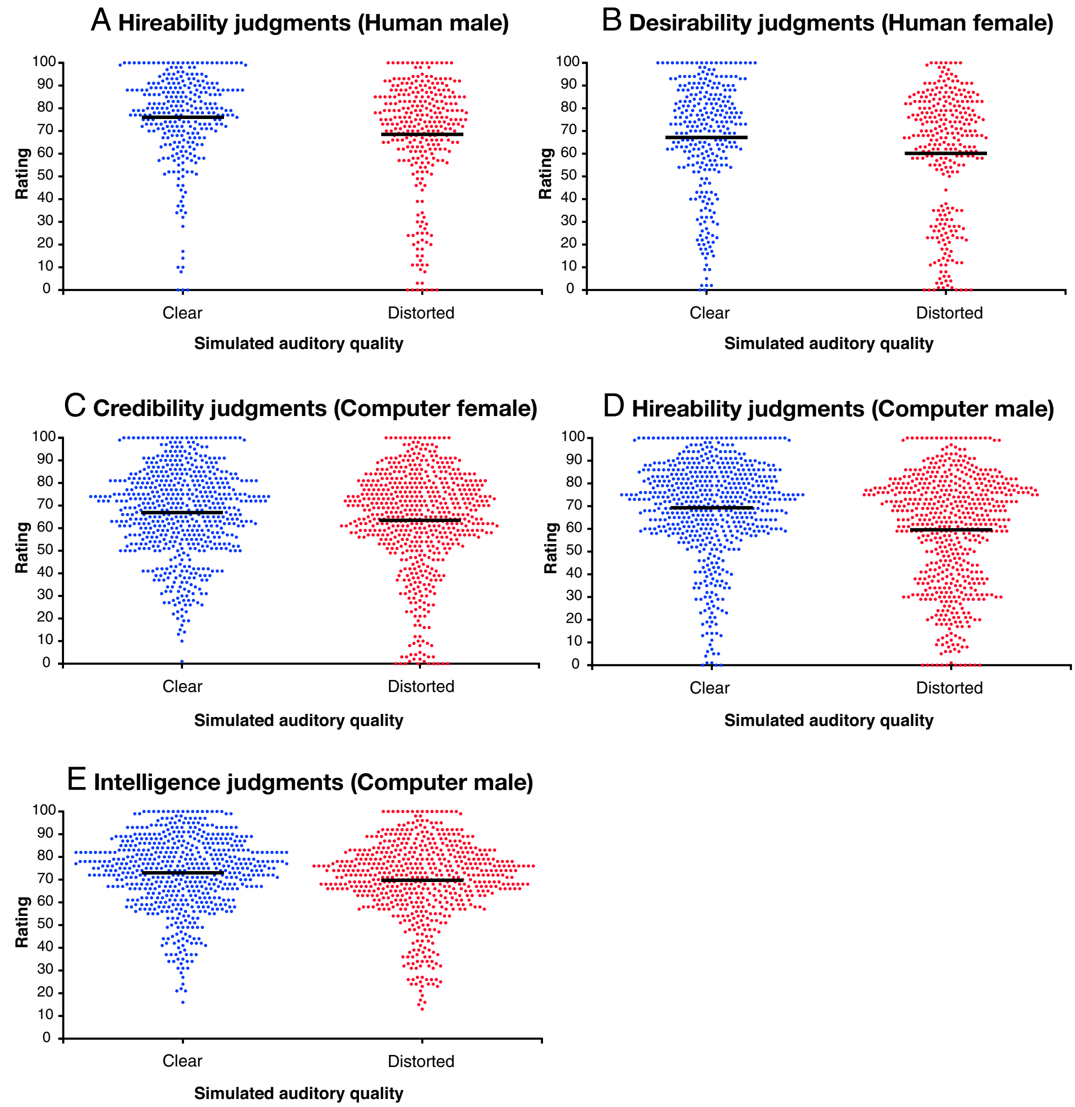

Fig. 1. All individual data points and group distributions, depicted in beeswarm plots. (*A*) Experiment 1 hireability ratings, (*B*) Experiment 2 romantic desirability ratings, (*C*) Experiment 3 credibility ratings, (*D*) Experiment 4 hireability ratings, and (*E*) Experiment 5 intelligence ratings.

depicted in Fig. 1*B*) that they were less likely to go on a date with the speaker compared to those who heard the Clear recording: 60.16 (SD = 27.48) vs. 67.13 (SD = 24.42), $t(598) = 3.28$, 95% CI [2.80, 11.14], $P = 0.001$, d = 0.27. Subsequent word-for-word transcription accuracy was again virtually identical across the two recordings: 99.30% (SD = 0.01) vs. 99.33 % (SD = 0.02), $t(598) = 0.19$, 95% CI [0.00, 0.00], $P = 0.851$, d = 0.02, $BF_{10} = 0.093$; coder agreement 98.24%.

Credibility. While the fluency effects in Experiments 1 and 2 controlled for brute understandability of individual words (via the transcription task), it is possible that important semantic information was still lost (e.g, perhaps the audio distortions in turn led to distortions of prosody that are intrinsic in human speakers). Further, it is possible that the recognition that a speaker is utilizing a low-quality microphone per se might give indications about various traits of the speaker (e.g., socioeconomic status, professionalism, tech-savviness). Thus for Experiments 3 to 5, the stimuli were composed of clearly synthetic computerized speech which minimizes both of these concerns (as the quality of computerized speech cannot be rationally attributed to the speaker themselves, and computerized speech lacks meaningful [or context-driven] suprasegmental cues, thus reducing the potential impact of diminished prosody in the Distorted conditions).

Beyond preference judgments, we also tested influences on an entirely different kind of judgment, involving credibility (Experiment 3). Listeners were asked to imagine that they worked at a car insurance company and had to make judgments about how much a driver was at fault in an accident and then heard the driver's statement. To help ensure that listeners could not possibly attribute the auditory quality to the driver themselves, the recording was played in a British female voice that was clearly computer-synthesized (Audios S5 and S6). When asked to judge how credible they found the speaker's statement on a continuous scale from "Very Unbelievable" to "Very Believable", listeners who heard the Distorted recording indicated (as depicted in Fig. 1*C*) that the speaker was less credible compared to those who heard the Clear recording: 63.46 (SD = 22.26) vs. 66.90 (SD = 19.44), $t(1198) = 2.85$, 95% CI [1.07, 5.80], $P = 0.004$, d = 0.16. Subsequent word-for-word transcription accuracy was again virtually identical across the two recordings, though the small difference was statistically reliable: 99.00% (SD = 0.02) vs. 99.50% (SD = 0.01), $t(1198) = 4.91$, 95% CI [0.00, 0.00], $P < 0.001$, d = 0.28; coder agreement 85.45%. As a result, we also showed that the influence of auditory quality on credibility judgments persisted even when considering only those listeners with perfect accuracy (Distorted n = 396, Clear n = 469). We randomly chose 396 judgments from Clear recordings with perfect transcriptions and compared the resulting judgments to those from the full set of Distorted recordings with perfect transcriptions. Averaged across 1000 random samples of the former, listeners who heard the Distorted recordings still judged the speaker to be less credible: 62.47 (SD = 0.00) vs. 66.31 (SD = 0.40), average $t(790) = 2.64$ (SD = 0.28), average $P = 0.011$ (SD = 0.01), average d = 0.19 (SD = 0.02).

Hireability (Computerized Voice Replication). To ensure that the results from Experiment 1 (involving hireability) did not depend on the human speaker's voice, we also replicated this effect with a computer-synthesized voice (Experiment 4)—using the same personal statement text, now recorded in an artificial American male voice (Audios S7 and S8). Listeners who heard the Distorted recording indicated (as depicted in Fig. 1*D*) that they were less likely to hire the speaker compared to those who heard the Clear recording: 59.57 (SD = 24.65) vs. 69.30 (SD = 20.88), $t(1198) = 7.38$, 95% CI [7.14, 12.32], $P < 0.001$, d = 0.43.

Intelligence. We generalized these effects to a more abstract trait of judged intelligence (Experiment 5), using the same materials and voice from Experiment 4. When asked to judge how intelligent they found the speaker on a continuous scale from "Very Unintelligent" to "Very Intelligent", listeners who heard the Distorted recording indicated (as depicted in Fig. 1*E*) that the speaker was less intelligent compared to those who heard the Clear recording: 69.70 (SD = 17.03) vs. 73.01 (SD = 15.95), $t(1198) = 3.48$, 95% CI [1.44, 5.18], $P < 0.001$, d = 0.20.

Hireability (Specificity Control). Might such effects reflect wider negative impressions, which are not specific to the speaker? This is possible, though in general disfluency can drive both negative impressions (e.g., leading to greater unpleasantness judgments) and positive impressions (e.g., leading to greater arousal and interest) (36, 37). To explore this experimentally—and in particular to ensure that these results reflect judgments of the speaker and not merely influences of disfluency on generalized affect or other preference judgments—we replicated Experiment 1 but also asked (in counterbalanced order) whether listeners themselves would enjoy working as a hiring manager. And listeners additionally completed the Positive and Negative Affect Schedule (PANAS) questionnaire (38), to assess their immediate affect. Listeners who heard the Distorted recording again indicated that they were less likely to hire the speaker compared to those who heard the Clear recording: 68.60 (SD = 19.90) vs. 74.59 (SD = 19.37), $t(598) = 3.74$, 95% CI [2.84, 9.14], $P < 0.001$, d = 0.31. However, there was no significant difference in the resulting prospective enjoyment ratings: 59.01 (SD = 26.48) vs. 58.04 (SD = 25.39), $t(598) = 0.46$, 95% CI [−5.13,3.19], $P = 0.65$, d = 0.04—with a significant interaction between these measures, $F(1196) = 6.87$, $P = 0.009$, $np2 = 0.006$. There was also no significant difference in resulting positive or negative affect scores: Positive 30.65 (SD = 8.61) vs. 30.86 (SD = 8.59), $t(598) = 0.29$, 95% CI [−1.17, 1.59]. $P = 0.77$, d = 0.02; Negative 13.11 (SD = 4.81) vs. 12.61 (SD = 4.52), $t(598) = 1.29$, 95% CI [−1.24, 0.26], $P = 0.20$, d = 0.11.

Discussion

Common disfluent auditory signals (as in tinny speech due to poor microphone quality) led to decreased judgments of intelligence, hireability, credibility, and romantic desirability. These effects seemed highly specific along two key dimensions. First, the results appeared to reflect (dis)fluency itself, rather than any difference in the resulting comprehensibility of the messages. (Indeed, this study is unique to our knowledge in its checks for word-for-word transcription accuracy in all subjects for Experiments 1 to 3.) Second, the influence of (dis)fluency seemed constrained to (a wide array of) judgments about the speakers, while having no appreciable effect (in Experiment 6) on more general judgments or affect.

Disfluent speech can come in many varieties, and it may be reasonable for some such factors to affect estimates of character traits or to influence concrete decisions such as hiring. For example, a speaker with a flat affect or who slurs their words might be less effective as a teacher or salesperson. But the manipulations of auditory fluency explored here lacked any such potential normative relevance, as they clearly reflected the superficial vehicle of speech, rather than intrinsic qualities of the speaker. The Distorted stimuli could not possibly be mimicked via a speaker's natural voice, and both the Clear and Distorted computer-synthesized voices were obviously not directly produced by the speakers themselves. Indeed, these manipulations were specifically designed to effectively "absolve" the speakers of responsibility for the disfluency—with even the instructions themselves conveyed in the same tinny speech in the disfluent condition (a manipulation for which the speaker clearly could not have been responsible). The effects reported here are thus inappropriate and irrational, making them especially important to identify.

The potential applied significance of these results seems equally relevant, if not more so. The kinds of distortions tested here are ubiquitous in everyday videoconferencing. (The initial idea for this study arose during a committee meeting that was occurring virtually. One faculty member was participating via high-quality equipment meant for music recording, while another seemed to be using an especially old/cheap computer microphone—and we noticed that the former's contributions seemed somehow more profound.) As such, we suspect that the effects reported here could really matter, particularly in scenarios where even minor decision-making biases can have significant consequences (e.g., when applying for a highly competitive job, or in high-stakes business negotiations). Moreover, these effects are potential sources of unintentional bias and discrimination, given the likelihood that microphone quality is correlated with socioeconomic status. [As of this writing, the top USB microphone recommendation from the New York Times, which "produced clear, rich-sounding recordings and preserved our speaker's natural vocal warmth", costs $110 (39)].

This real-world context might also be theoretically unique vis-à-vis fluency studies, in terms of speakers' potential knowledge of the relevant variable in the first place. If you write a document in a disfluent font, for example, you know what font you are using. (You can see it too.) If you speak in a mumbling or slurred tone in real-life interactions, you know to some degree what you sound like. (You can hear yourself too.) And even in the context of the video signal during videoconferencing, you know what you look like to others. (You can see the same camera image that they do.) But auditory fluency based on microphone quality during videoconferencing is completely different: Here, you (and you alone!) cannot hear what you sound like to others. So, before joining your next videoconference, you may want to consider how much a cheap microphone may really be costing you.

Materials and Methods

Preregistrations. All experiments were preregistered before data collection began and can be viewed at https://aspredicted.org/TBJ_MYA (Experiment 1), https://aspredicted.org/WCA_EIH (Experiment 2), https://aspredicted.org/HNE_LZQ

(Experiment 3), https://aspredicted.org/LEU_MSE (Experiment 4), https://aspredicted.org/XSM_EVN (Experiment 5), and https://aspredicted.org/dq9y-r8ym.pdf (Experiment 6).

Experiment 1: Judgments of Hireability (Human Male Voice).

Participants. A total of 600 unique listeners (300 for each of the two conditions described below) were recruited from the Prolific online platform and participated for monetary compensation. Prescreening criteria required listeners to have English as a first language, a Prolific approval rating of at least 95%, previous completion of at least 100 Prolific studies, no prior participation in another experiment from this project, and the use of a laptop or desktop computer (but not a phone or tablet). This sample size was preregistered before data collection began. At the end of the experiment, listeners completed a debriefing questionnaire, which included questions asking about how well they paid attention (on a scale of 1 to 100) and if they experienced any technical difficulties or distractions while playing the audio. All experimental methods and procedures were approved by the Yale University Institutional Review Board, and all listeners confirmed that they had read and understood a consent form outlining their risks, benefits, compensation, and confidentiality and that they agreed to participate in the experiment.

Per the preregistered exclusion criteria, 48 listeners (7.4%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) Self-reported attention ratings below 70/100 ($n = 12$), 2) technical difficulties or distractions ($n = 3$), 3) response times more than 2.5 SD from the mean ($n = 14$), and 4) transcription accuracy scores more than 2 SD from the mean ($n = 19$). (In addition, 12 other listeners were excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.) In practice, including the excluded listeners (where their data was recorded)—whether here or in all the other experiments—had no impact on the qualitative patterns reported in the main text.

Apparatus. All experiments were completed on custom webpages created with software written in a combination of PHP, JavaScript, CSS, and HTML, with the jsPsych libraries (40). Before beginning each experiment, listeners were asked to either wear headphones or move to a quiet environment. To discourage multitasking, listeners completed the experiment with their browser in full-screen mode. During debriefing, listeners reported whether they heard the stimuli using speakers or headphones; because later analyses revealed no differences in the reported effects based on such hardware, we collapsed over this factor in all the reported analyses.

Stimuli. The Clear audio stimulus (in mp3 format) was a recorded narration by a naive male speaker. The recording (Audio S1) was 27.5 s and consisted of the following spoken text: "After 8 y in sales, I am currently seeking a new challenge which will utilize my meticulous attention to detail, and friendly, professional manner. I am an excellent fit for your company and will be an asset to your team as a senior sales manager. As an experienced sales manager with my previous company, my tenacious and proactive approach resulted in numerous important contract wins. Through this experience, I have improved and developed my networking skills, which have proven to be very effective in increasing my number of sales." (41). The Distorted audio stimulus (Audio S2) was created by modifying the Clear recording using the open-source VST effect "MDACombo" (42) in the TwistedWave Online Audio Editor (https://twistedwave.com/online), with the following settings: Model: 4x12>; Drive: −55; Bias: 83; Output: 0; Process HPF Frequency: 20%; HPF Resolution: 90%. This resulted in speech that was fully comprehensible, but that had a high-frequency tinny quality like that commonly experienced during videoconferences with a low-quality computer microphone. To minimize the salience of the distortion, audio played during the instructions (e.g., while testing the volume) when using this stimulus was also distorted in the same way. Stimuli for the word-for-word transcriptions were created by breaking up each of the Clear and Distorted recordings into 16 short segments (each approximately 2 s long, always cut at the same timestamps in both conditions).

Procedure and design. Listeners were randomly assigned to one of two conditions: half of the listeners heard Clear recordings, and the other half heard Distorted recordings. Listeners first heard a short audio recording which confirmed that they could hear speech, and which allowed them to adjust the volume to a comfortable level. (Listeners assigned to the Unclear condition heard these instructions with the same distortion manipulation used for the subsequent audio stimuli.) Listeners were then presented with the following centered written prompts [across a few separate screens, with all written text presented in black "Open Sans" font, scaled to a point size that was 1.5% of the full-screen pixel width of the listener's display, on a light gray (#D3D3D3) background]: "For this study, we'd like you to imagine that as part of your job, you are tasked with making a hiring decision for a highly competitive position: senior sales manager. You will listen to a few lines from a personal statement from an application for this position. We will then ask you a few questions about your impressions of the personal statement and its author. Click on the 'Play' button below to play the audio clip of the personal statement. You will only be able to listen to this once. (You cannot replay it.)" Listeners then heard the audio stimuli after clicking on the relevant button (which then disappeared). After the recording finished playing, listeners were asked "What is the likelihood that you would hire this person?", and responded by using their computer mouse to position a slider (a light blue circle 35px in diameter) on a continuous scale (depicted by a white bar, 25px high, with a width equal to 75% of the full-screen width) from Very Unlikely to Very Likely (with these terms appearing just below the bar on the far left and right, respectively). (Listeners could adjust the marker's position as many times as they wished, after which they clicked on a "Continue" button.) Responses were recorded as values ranging from 0 (matching the bar's far left) to 100 (representing the bar's far right). Listeners next completed a transcription task where they were played the same recorded personal statement, except now divided into smaller sections as described above. Listeners pressed a button to play each section and could only play each section once. After each section, listeners simply transcribed the section they heard into a text box (with a width equal to 60% of the full screen's height), after which they could play the next section. If listeners did not enter any text into the box, an error message appeared saying, "You didn't enter any text. Please type as much as you can remember and click "Next"."

Data analysis. The dependent measure was the "hireability" rating listeners gave after hearing the narrated personal statement. A between-samples two-tailed *t* test was used to determine whether the average ratings differed significantly between the Clear and Distorted conditions. We also used a between-samples two-tailed *t* test to determine whether the transcription accuracy significantly differed between conditions.

Word-for-word transcriptions were first stripped of punctuation and capitalization and then processed through a custom automated pattern matching algorithm to mark error-free transcription intervals as fully accurate. Remaining transcription intervals were then manually checked for errors by two coders (one naive) who were blind to the condition, each of whom viewed the flagged responses for each interval in turn (with the responses across the listeners from both conditions fully randomized within each interval, separately for each coder). Extra words, missing words, or substituted words each constituted a single error, but incorrect spelling or concatenations (e.g., "can't" vs "can not") were not counted as errors. Each listener's accuracy % was scored as (the total number of words in the actual text—the total number of errors)/the total number of words.

Experiment 2: Judgments of Desirability (Human Female Voice). This experiment was identical to Experiment 1 except as noted. First, 600 new listeners were recruited, with this preregistered sample size chosen to match that of Experiment 1. Per the preregistered exclusion criteria, 38 listeners (6.0%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) self-reported attention ratings below 70/100 ($n = 10$), 2) technical difficulties or distractions ($n = 3$), and 3) response times more than 2.5 SD from the mean ($n = 8$), and 4) transcription accuracy scores more than 2 SD from the mean ($n = 17$). (In addition, 8 other listeners were excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.)

The audio stimulus (Audios S3 and S4) was a recorded narration by a naive female speaker. The recording was 29.5 s long and consisted of the following spoken text: "I really enjoyed reading your profile. Most people on these dating sites come off as fake or desperate, but you seem completely genuine! I 'll be honest, I wasn't interested in most of the people here, however, you are an exception. We clearly have common interests and I think that is extremely important when getting to know someone. In the past, I've suffered through terrible dates with people I didn't click with, but I have a feeling that won't be a problem with us. I hope you feel the same way!"

Listeners were first presented with the following prompt: "For this study, please imagine you are single and have signed up for an online dating service where you virtually speed-date in order to meet women: You meet multiple women one after the other, and they only have around 30 s to tell you about themselves. You will now listen to a recording from one person who is hoping to get a date with you, and we will then ask you about what you heard." After listening to the recordings, listeners were asked "How likely would you be to go on a date with this person?".

Experiment 3: Judgments of Credibility (Computer Female Voice). This experiment was identical to Experiment 1 except as noted. First, 1,200 new listeners were recruited, with this sample size chosen to be exactly double that of Experiment 1 (with this preregistered decision made prior to data collection due to the possibility that computerized speech would be generally less fluent). Per the preregistered exclusion criteria, 50 listeners (4.0%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) self-reported attention ratings below 70/100 (n = 13), 2) technical difficulties or distractions (n = 9), 3) response times more than 2.5 SD from the mean (n = 10), and 4) transcription accuracy scores more than 2 SD from the mean (n = 18). (In addition, 32 other listeners were excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.)

The audio stimulus (Audios S5 and S6) was generated by Amazon Polly text-to-speech software (https://aws.amazon.com/polly/) using a simulated female British English voice ("Amy"). The recording was 25 s long and consisted of the following spoken text: "I was backing out of a parking space when I saw another car was also backing up. Right before they hit me, my brakes were fully applied and I was completely stopped. I'm not sure what the other driver is saying, I just know they were the only one moving, not me. The other driver clearly caused this accident. Maybe they were on their mobile phone or distracted by something. But either way, I can honestly say I did not cause this accident." Stimuli for the word-for-word transcriptions were created by breaking up each of the Clear and Distorted recordings into 17 short segments (each approximately 2 s long, always cut at the same timestamps in both conditions).

Listeners were first presented with the following prompt: "For this study, I'd like you to imagine that you work at a car insurance company, and your job is to decide how much a driver is at fault in a car accident. (Many factors will ultimately be involved in the company's decision, but your job is just to make a subjective judgment based on what you hear.) You are currently investigating a minor accident in a parking lot and both drivers claim they are not at fault. You will now listen to a brief statement from one of the drivers about what happened, and we will ask you questions about what you heard." After listening to the recordings, listeners were asked "How believable did you find the person's statement?" and responded with a continuous scale from Very Unbelievable to Very Believable.

Experiment 4: Judgments of Hireability (Computer Male Voice). This experiment was identical to Experiment 1 except as noted. First, 1,200 new listeners were recruited, with this preregistered sample size chosen to match that of Experiment 3. Per the preregistered exclusion criteria, 47 listeners (3.8%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) self-reported attention ratings below 70/100 (n = 27), 2) technical difficulties or distractions (n = 1), and 3) response times more than 2.5 SD from the mean (n = 19). (In addition, 1 listener was excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.) The audio stimulus (Audios S7 and S8) was generated (as in Experiment 3) using a simulated male American English voice ("Matthew"). The recording was 28.5 s long and consisted of the same text as used in Experiment 1.

Experiment 5: Judgments of Intelligence (Computer Male Voice). This experiment was identical to Experiment 4 except as noted. First, 1,200 new listeners were recruited, with this preregistered sample size chosen to match that of Experiment 3. Per the preregistered exclusion criteria, 54 listeners (4.3%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) self-reported attention ratings below 70/100 (n = 33), 2) technical difficulties or distractions (n = 2), and 3) response times more than 2.5 SD from the mean (n = 19). (In addition, 9 listeners were excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.) After listening to the recordings (identical to those in Experiment 4), listeners were asked "How intelligent is the author of this personal statement?" and responded with a continuous scale from Very Unintelligent to Very Intelligent.

Experiment 6: Specificity Control on Judgments of Hireability (Human Male Voice). This experiment was identical to Experiment 1 except as noted. First, 600 new listeners were recruited, with this preregistered sample size chosen to match that of Experiment 1. Per the preregistered exclusion criteria, 29 listeners (4.6%) were excluded from the analyses (with replacement) for the following criteria, in the following order: 1) self-reported attention ratings below 70/100 (n = 0), 2) technical difficulties or distractions (n = 0), and 3) response times more than 2.5 SD from the mean (n = 11). (In addition, 18 other listeners were excluded because they completed the experiment after the preregistered postexclusions sample size had been reached.)

In addition to the "hireabilty" question from Experiment 1, subjects were also asked (in counterbalanced order) a new question: "Overall, do you think you'd find working as a hiring manager to be enjoyable or unenjoyable?", responding with a continuous scale from Very Unenjoyable to Very Enjoyable. Afterward, instead of the transcription task, subjects completed a 20-question PANAS questionnaire (38) to determine their immediate positive and negative affect.

Data, Materials, and Software Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Stefan Uddenberg and Adrienne Walter-Terrill for recordings, Vivian Wang for manual transcription coding, and Nick Turk-Browne and Greg McCarthy for inspiring the study with their microphones. For helpful comments, we thank Clara Colombatto, and other members of the Yale Perception and Cognition Laboratory.

1. S. E. Asch, Forming impressions of personality. *J. Abnorm. Soc. Psychol.* **41**, 258–290 (1946), 10.1037/h0055756.
2. N. Ambady, The perils of pondering: Intuition and thin slice judgments. *Psychol. Inq.* **21**, 271–278 (2010), 10.1080/1047840x.2010.524882.
3. N. Ambady, F. J. Bernieri, J. A. Richeson, Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Adv. Exp. Soc. Psychol.* **32**, 201–271 (2000), 10.1016/S0065-2601(00)80006-4.
4. P. Borkenau, N. Mauer, R. Riemann, F. M. Spinath, A. Angleitner, Thin slices of behavior as cues of personality and intelligence. *J. Pers. Soc. Psychol.* **86**, 599–614 (2004), 10.1037/0022-3514.86.4.599.
5. N. Ambady, R. Rosenthal, Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J. Pers. Soc. Psychol.* **64**, 431–441 (1993), 10.1037/0022-3514.64.3.431.
6. M. W. Kraus, B. Torrez, J. W. Park, F. Ghayebi, Evidence for the reproduction of social class in brief speech. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22998–23003 (2019), 10.1073/pnas.1900500116.
7. M. O'Sullivan, P. Ekman, W. Friesen, K. R. Scherer, What you say and how you say it: The contribution of speech content and voice quality to judgments of others. *J. Pers. Soc. Psychol.* **48**, 54–62 (1985), 10.1037/0022-3514.48.1.54.
8. J. J. M. O'Connor, P. Barclay, The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evol. Hum. Behav.* **38**, 506–512 (2017), 10.1016/j.evolhumbehav.2017.03.001.
9. A. Oleszkiewicz, K. Pisanski, K. Lachowicz-Tabaczek, A. Sorokowska, Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychon. Bull. Rev.* **24**, 856–862 (2017), 10.3758/s13423-016-1146-y.
10. M. Mileva, N. Lavan, Trait impressions from voices are formed rapidly within 400 ms of exposure. *J. Exp. Psycho. Gen.* **152**, 1539–1550 (2023), 10.1037/xge0001325.
11. W. T. Fitch, J. Giedd, Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am.* **106**, 1511–1522 (1999), 10.1121/1.427148.
12. W. T. Fitch, "Production of vocalizations in mammals" in *Encyclopedia of Language Linguistics*, K. Brown, Ed. (Elsevier Science, 2006), pp. 115–121. 10.1016/B0-08-044854-2/00821-X.
13. A. Sell *et al.*, Adaptations in humans for assessing physical strength from the voice. *Proc. R. Soc. Lond. B.* **277**, 3509–3518 (2010), 10.1098/rspb.2010.0769.
14. J. Stern *et al.*, Do voices carry valid information about a speaker's personality?. *J. Res. Pers.* **92**, 104092 (2021), 10.1016/j.jrp.2021.104092.
15. C. A. Klofstad, Candidate voice pitch influences election outcomes. *Polit. Psychol.* **37**, 725–738 (2015), 10.1111/pops.12280.
16. D. M. Oppenheimer, Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Appl. Cogn. Psychol.* **20**, 139–156 (2006), 10.1037/e633872013-573.
17. R. K. Guenther, Does the processing fluency of a syllabus affect the forecasted grade and course difficulty? *Psychol. Rep.* **110**, 946–954 (2012), 10.2466/01.11.28.PR0.110.3.946-954.
18. V. P. Magnini, S. Kim, The influences of restaurant menu font style, background color, and physical weight on consumers' perceptions. *Ing. J. Hosp. Manag.* **53**, 42–48 (2016), 10.1016/j.ijhm.2015.11.001.
19. A. L. Alter, D. M. Oppenheimer, Uniting the tribes of fluency to form a metacognitive nation. *Pers. Soc. Psychol.* **13**, 219–235 (2009), 10.1177/1088868309341564.
20. D. M. Oppenheimer, The secret life of fluency. *Trends Cogn. Sci.* **12**, 237–241 (2008), 10.1016/j.tics.2008.02.014.
21. S. Topolinski, "The sources of fluency: Identifying the underlying mechanisms of fluency effects" in *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour*, C. Unkelbach, R. Greifender, Eds. (Psychology Press, 2013), pp. 33–49.

22. N. Schwarz, M. C. Jalbert, T. Noah, L. Zhang, Metacognitive experiences as information: Fluency in consumer judgment and decision making. *Consum. Psych. Rev.* **4**, 4–25 (2021), 10.1002/arcp.1067.
23. S. M. Herzog, R. Hertwig, "The ecological validity of fluency" in *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour*, C. Unkelbach, R. Greifender, Eds. (Psychology Press, 2013), pp. 190–219.
24. R. Greifeneder, H. Bless, S. G. Scholl, "About swift defaults and sophisticated safety nets: A process perspective on fluency's validity" in *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour*, C. Unkelbach, R. Greifender, Eds. (Psychology Press, 2013), pp. 220–233.
25. C. Unkelbach, R. Greifeneder, "A general model of fluency effects in judgment and decision making" in *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour*, C. Unkelbach, R. Greifender, Eds. (Psychology Press, 2013), pp. 11–32.
26. J. G. Beerends, F. E. De Caluwe, The influence of video quality on perceived audio quality and vice versa. *J. Audio Eng. Soc.* **47**, 355–362 (1999).
27. S. K. Carpenter, M. M. Wilford, N. Kornell, K. M. Mullaney, Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychon. Bull. Rev.* **20**, 1350–1356 (2013), 10.3758/s13423-013-0442-z.
28. K. Schoenenberg, A. Raake, J. Koeppe, Why are you so slow? – Misattribution of transmission delay to attributes of the conversation partner at the far-end *Int. J. Hum-Comput. Int.* **72**, 477–487 (2014), 10.1016/j.ijhcs.2014.02.004.
29. E. J. Newman, N. Schwarz, Good sound, good research: How audio quality influences perceptions of the researcher and research. *Sci. Commun.* **40**, 246–257 (2018), 10.1177/1075547018759345.
30. E. Bild *et al.*, Sound and credibility in the virtual court: Low audio quality leads to less favourable evaluations of witnesses and lower weighting of evidence. *Law Hum. Behav.* **45**, 481–495 (2021), 10.1037/lhb0000466.
31. J. F. Dovidio, A. Gluszek, "Accents, nonverbal behavior, and intergroup bias" in *The Handbook of Intergroup Communication*,H. Giles, Ed. (Routledge, 2012), pp. 87–99.
32. M. Dragojevic, H. Giles, I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Hum. Commun. Res.* **42**, 396–420 (2016), 10.1111/hcre.12079.
33. M. Dragojevic, H. Giles, A. C. Beck, N. T. Tatum, The fluency principle: Why foreign accent strength negatively biases language attitudes. *Commun. Monogr.* **84**, 385–405 (2017), 10.1080/03637751.2017.1322213.
34. S. Lev-Ari, B. Keysar, Why don't we believe non-native speakers? The influence of accent on credibility *J. Exp. Soc. Psychol.* **46**, 1093–1096 (2010), 10.1016/j.jesp.2010.05.025.
35. J. N. Fuertes, W. H. Gottdiener, H. Martin, T. C. Gilbert, H. Giles, A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *Eur. J. Soc. Psychol.* **42**, 120–133 (2012), 10.1002/ejsp.862.
36. A. A. Labroo, A. Pocheptsova, Metacognition and consumer judgment: Fluency is pleasant but disfluency ignites interest. *Curr. Opin. Psychol.* **10**, 154–159 (2016), 10.1016/j.copsyc.2016.01.008.
37. H. Song, N. Schwarz, If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psych. Sci.* **20**, 135–138 (2009), 10.1111/j.1467-9280.2009.02267.x.
38. D. Watson, L. A. Clark, A. Tellegen, Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988), 10.1037//0022-3514.54.6.1063.
39. M. Pinola, The best USB microphone. *The New York Times Wirecutter* (9 August 2023). https://web.archive.org/web/20231207203021/https://www.nytimes.com/wirecutter/reviews/the-best-usb-microphone/. Accessed 29 July 2024.
40. J. R. de Leeuw, jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12 (2015), 10.3758/s13428-014-0458-y.
41. J. Reed, *The 7-Second CV: How to Land the Interview* (Virgin Books, 2019).
42. K. Kellett, *VST MDACombo* (1998). https://web.archive.org/web/20211225205533/http://mda.smartelectronix.com/. Accessed 29 July 2024.