When No One Hears the Recording: Typed Versus Spoken Input in App-Processed 360-Degree Feedback

Stated preferences of 56 professionals. Preprint v1 (exploratory interview study)

Alessio Biancheri[1]

8 August 2026

Status: Work in progress. Open research agenda in Section 10. Collaboration invited.

Abstract

AI-native multi-source feedback platforms can accept speech as raw input: the employee records their feedback into the app, the system transcribes, analyses and consolidates it, and the assessed person receives a written report. No individual recording is ever delivered to anyone. This severs the usual link between input modality and delivery modality. The standard advice that feedback is best delivered rich, ideally as a conversation, is advice about delivery, and it is silent about a recording that no human will ever receive. What employees actually prefer under these conditions is, to our knowledge, unstudied in the 360-degree feedback context.

We asked 56 working professionals a single standardized scenario question: your company runs a 360-degree campaign about a colleague you work with; you can type your feedback or record it as audio; the audio is only processed by the app, and the assessed person receives a consolidated written document, never your recording. Every respondent received the same wording, the same clarification, and one minute to think. **46 of 56 (82%) chose writing; 10 (18%) chose audio.**

The reasons are the finding. Writing-choosers cited control over wording and revision before submission, time to think, discretion in shared workspaces, discomfort with their own recorded voice, and unwillingness to leave voice recordings (identifiable, quasi-biometric data) in an employer's system. Audio-choosers cited speed, typing burden and more natural expression, several reasoning explicitly that "the app processes it anyway." Almost none of the reluctance referenced the removed listener. **Removing the human listener removes almost none of the resistance to speaking**: the barriers sit in the capture moment and in the stored artifact, not in the delivery.

The split replicates survey-methodology findings in a new domain. Offered a choice on narrative open-ended questions, roughly 78–83% of respondents answered by text in a recent experiment, and 94% chose text in an earlier panel study. A theme-by-theme comparison finds no interview theme that contradicts that literature, while two of our highest-ranked themes, self-presentation under assessment stakes and aversion to one's own recorded voice, do not appear in its catalogue of reasons at all. The same literature finds spoken answers longer, more spontaneous and richer in personal experience: the modality that yields the best raw material for a processing pipeline is the one four out of five respondents avoid. We state design implications: text as default, voice as an option, visible transcription with editing before submission, and deletion of raw audio by default.

We are explicit about what this study is and is not. It is a single-question, stated-preference exploration with a convenience sample, coded by a single rater who is also a vendor in this product category. It establishes that the preference exists, that it is strong, and what respondents say drives it. It does not establish behavior in a live campaign, and we specify the study that would (Section 10).

[1] ignostiq (Alesserg Technology OÜ), Tallinn, Estonia. a.biancheri@ignostiq.com. ORCID 0009-0009-9653-7091

Keywords: 360-degree feedback; multi-source feedback; voice input; speech interfaces; modality choice; open-ended questions; media synchronicity; self-presentation; voice privacy; HR technology

---

1. Introduction

The advice attached to workplace feedback for decades is to deliver it rich. Media richness theory holds that equivocal, socially delicate messages need rich channels, ideally face-to-face, where tone, immediacy and repair are available [Daft & Lengel, 1986]. Recent experimental work adds that hearing a voice humanizes the communicator in ways text does not [Kumar & Epley, 2021]. The practitioner instinct follows: if the feedback matters, have the conversation. The multi-source feedback literature, for its part, shows that whether 360-degree programmes change behavior depends heavily on process design [Bracken & Rose, 2011], and that feedback interventions in general are far from uniformly positive [Kluger & DeNisi, 1996].

An AI-native 360-degree platform changes the object this advice was written about. In such a system, an employee's spoken audio is an **input, not a message**. The pipeline is: record, transcribe, analyse, consolidate, and deliver as a written document. The assessed person receives themes and evidence aggregated across assessors. They never receive, and never hear, any individual recording. Nothing is delivered by voice; voice is merely one of two ways to get raw material into the system.

Which way should intuition break here? There is a respectable argument that audio should become *more* attractive once no human hears it: the speaker keeps all the ease and richness of talking, while the exposure that makes voice messages risky (a colleague judging your tone, recognizing your voice in a nominally anonymous process, replaying your worst five seconds) is removed by construction. If reluctance to record feedback were mainly about being heard by the colleague, removing the colleague should remove the reluctance.

It does not. That is the empirical point of this paper. Asked to choose, 46 of 56 professionals (82%) chose typing, and their stated reasons barely reference the recipient at all. The reasons live in the capture moment (the room you would have to find, the voice you would have to hear back, the sentence you cannot un-say) and in the stored artifact: a recording of your voice, sitting in an employer's system. The conversation intuition locates the cost of speaking in the listening. Our respondents locate it in the recording.

1.1 Contributions

1. **The audio-as-message versus audio-as-input distinction**, stated explicitly for feedback systems. The two objects have different risk profiles, and intuitions about one do not transfer to the other (§3).
2. **A 56-person stated-preference result with reasons**, collected under explicit no-listener conditions: 82% writing / 18% audio, with a coded account of why (§5).
3. **A theme-by-theme comparison with the survey-methodology literature**: five themes confirm findings that literature established with stronger designs, two are effectively new to the assessment context, none contradicts it, and we list the questions neither literature answers (§6).
4. **The trade-off between adoption and information** for feedback pipelines: the avoided modality produces the richer raw material. Seven design principles follow (§7–8).
5. **An honest account of what this study cannot support**, and the controlled study that would settle the question (§9–10).

---

2. Related work

The delivery intuition. Kluger and DeNisi [1996] established that feedback interventions have highly variable effects, including negative ones; Bracken and Rose [2011] argue that process and instrument design determine whether multi-source feedback changes behavior. Media richness theory [Daft & Lengel, 1986] and the modern media-psychology result that voice humanizes its speaker [Kumar & Epley, 2021] both concern what a **human recipient** experiences. In our scenario there is no human at the audio's other end; the framework is not wrong, it is simply about a different object.

Text versus voice as input to a system. Survey methodology has run almost exactly this comparison for a decade, because web surveys can offer open-ended questions with typed or spoken answers. The pattern is consistent. Voice input creates friction and nonresponse [Revilla, Couper, Bosch & Asensio, 2020]; a majority of panelists say they are unwilling to answer by voice, citing preference for written communication and lack of habit [Lenzner & Höhne, 2022]. When respondents are given the choice, text dominates: 94% chose text in a probability-based panel [Meitinger, van der Sluis & Schonlau, 2022, as summarized in Revilla & Couper, 2026], and in a recent experiment offering text, dictation and voice recording, only about 17–22% used any voice input, with response rates significantly lower when voice was pushed [Revilla & Couper, 2026]. Asked why they avoid voice, respondents cite other people being present, difficulty expressing ideas orally, and doubts about what happens to the recording; trust in confidentiality predicts uptake [Revilla & Couper, 2024].

The counterpoint is just as consistent: spoken answers are longer, more spontaneous, lexically simpler and richer in personal experience than typed ones [Gavras, Höhne, Blom & Schoen, 2022; Höhne, Gavras & Claassen, 2024; Galasso, Nannicini & Nozza, 2024]. Voice yields richer data, from fewer people, with higher break-off.

Mechanisms. Media synchronicity theory names the two properties our respondents keep reaching for: *rehearsability* (fine-tune the message before it goes) and *reprocessability* (re-read and reconsider it) [Dennis, Fuller & Valacich, 2008]. Walther's work on selective self-presentation explains why editable, asynchronous text is attractive precisely when the message is identity-relevant [Walther, 2007]. The strangeness of one's own recorded voice has a physiological basis: we normally hear ourselves partly through bone conduction, so the recording genuinely is a different voice [Orepic, Kannape, Faivre & Blanke, 2023]. And voice is behavioral biometric data, identifiable and rich in secondary inferences, which makes storing raw audio in an employer system a real privacy question, not a neurotic one [Hanisch, Arias-Cabarcos, Parra-Arnau & Strufe, 2025].

Our position. We contribute no new mechanism. What we add is a test of whether these input-side preferences survive in a setting the survey literature does not cover, evaluating an identifiable colleague inside one's own company, and an explicit statement of the input/message distinction for feedback systems. Survey questions have no recipient at all; a 360 campaign has a very salient one. If preferences match across the two settings, the mechanisms that matter are capture-side, not recipient-side. They match.

---

3. Audio as message versus audio as input

Speaking feedback into a phone can produce two different objects, depending on the system behind the microphone.

> **Audio as message.** The recording itself is delivered. The recipient hears voice, tone, hesitation and identity. Voicemail, voice notes, and a recorded feedback message all work this way.

Audio as input. The recording is raw material for a pipeline. It is transcribed and analysed; what any human eventually receives is a derived document. The recording is heard by no one, by design.

The distinction matters because the two objects have different risk surfaces:

Property	Audio as message	Audio as input (this scenario)
Who hears the recording	The recipient	No human; the system processes it
What reaches the assessed person	Voice, tone, wording, identity	Consolidated written themes
Tone risk	Heard directly by the recipient	Reduced to extracted content
Voice identifiability	Recipient recognizes the voice	Recording stored in the system; output aggregated
Revision	Re-record	Re-record; or edit a transcript, if offered
Persistence	Recipient may keep the audio	Employer system keeps it, per retention policy

Each frame generates a prediction. If people reason in the **message frame**, reluctance to record should be driven by the recipient (being judged, being recognized) and should collapse when the recipient is removed. If people reason in the **input frame**, reluctance should be driven by capture (environment, self-consciousness, loss of revision) and by storage (an identifiable recording held by the employer), and should survive the removal of the listener intact. The interviews discriminate between these predictions.

---

4. Method

4.1 Question and procedure

Each respondent was asked one standardized scenario question, identical for all 56 (full wording in Appendix A): your company is running a 360-degree feedback campaign; you are asked to give feedback about a colleague you work with; you can either write your feedback or record it as audio into the app. It was stated explicitly that the audio is only processed by the app and that the assessed person receives a consolidated written document, never the recording. Respondents were then given **one minute to think**, and answered with a choice and their reasons. The same wording, the same clarification and the same procedure were used for every respondent.

Interviews were conducted individually by the author, conversationally, without a further scripted protocol: one question, one minute, one answer, with natural follow-up where the respondent kept talking.

4.2 Sample

56 working professionals from the author's extended professional network: a convenience sample spanning varied roles and sectors, weighted toward office and knowledge work. Demographics were not systematically recorded, which we report as a limitation rather than disguise (§9).

4.3 Records and coding

Interviews were audio-recorded. The author reviewed all 56 recordings and extracted, for each respondent, the stated choice and the reasons given. Responses reported in this paper are **condensed paraphrases of the recorded answers, not full verbatim transcripts**, and are attributed by respondent code (P001–P056).

Each response was then assigned one **primary theme**, the reason the respondent led with or stressed, using the coding frame in Appendix B. Most answers contained more than one motive; primary-theme counts therefore flatten mixed motivations and should be read as coarse. **The coding was performed by a single rater, the author, retrospectively, and not blind.** This conditions everything that follows and we return to it in Section 9.

4.4 Ethics

Respondents spoke in a personal capacity about a hypothetical scenario. No employer commissioned the exercise, no employer data was involved, and no names, employers or demographics are reported. Quotes are paraphrased.

---

5. Results

5.1 The choice

Choice	Respondents	Share
Writing	46	82.1%
Audio	10	17.9%
Total	56	100%

Four out of five chose typing, in a scenario explicitly constructed so that no human would ever hear the recording.

5.2 Why writing (46 respondents)

Primary theme	Respondents	Share of 56
Editability / control	8	14.3%
Reflection / structure	7	12.5%
Voice / recording discomfort	7	12.5%
Psychological safety / self-presentation	7	12.5%
Privacy / environment	6	10.7%
Voice / data privacy	5	8.9%
Reviewability / clarity	4	7.1%
Habit / accessibility	2	3.6%

Quotes below are condensed paraphrases from the recordings (§4.3). The eight themes cluster into five stories.

Control of the message before it leaves. The largest cluster, editability (8), reflection (7) and reviewability (4), together 19 of 46, is media synchronicity theory spoken in plain language: rehearsability and reprocessability [Dennis et al., 2008]. *"Writing is easier to review and change. With audio, what do I do if I want to correct one sentence? Record the whole thing again?"* (P003). *"I can read it once as 'me' and once as the person receiving it. I usually change something on the second pass"* (P009). *"Sometimes my first reaction is not the fairest one. Writing slows me down enough to separate the example from the emotion"* (P010). *"I hate silences, so when I record I start filling them with whatever comes to mind"* (P001).

Self-presentation and emotional leakage. Seven respondents chose writing to control how they come across: Walther's selective self-presentation [2007] under assessment conditions. *"I can be*

honest without sounding impulsive. I don't want a bad five seconds of tone to become the story" (P018). *"When I'm annoyed with someone, you can hear it immediately in my voice. Text lets me give the useful part and remove the heat"* (P028). Note what is being managed here: not what the colleague will hear (they will hear nothing) but what the recording will contain.

The relationship with one's own recorded voice. Seven more located the problem in the recording experience itself: dislike of the recorded self-voice and refusal to replay it for quality control [Orepic et al., 2023], performance anxiety, accent self-consciousness. *"I don't really like the tone of my recorded voice. And if I don't like listening back to it, I'm definitely not going to review the recording properly before sending it"* (P002). *"Recording makes me perform. I start paying attention to myself instead of the feedback"* (P012).

The capture environment. Six respondents answered as workers in shared offices, not as media theorists. *"Typing is discreet. Saying someone's name and talking about their behaviour out loud is not"* (P013). *"I can type discreetly on a train or in an open office; I can't record discreetly"* (P041). Recording turns a private task public, and it adds logistics to a five-minute obligation: find a room, close a door.

What the company stores. Five respondents objected to the artifact: a voice recording, identifiable by nature, held in an employer system [Hanisch et al., 2025]. *"My voice is private. I wouldn't want my company storing recordings of me when the same feedback can be given in text"* (P008). *"If the feedback is supposed to be anonymous, audio immediately feels wrong. People recognise voices; my voice is basically my name"* (P049). *"Why collect something more identifying than necessary for a feedback exercise?"* (P050). *"I don't know what happens to the recording after I submit it. Text feels less invasive because at least I know exactly what I handed over"* (P033).

The remaining two cited habit and language: written workflows are the workplace norm (P034), and a non-native speaker can check a word in text where speech would force simplification (P046).

5.3 Why audio (10 respondents)

Primary theme	Respondents	Share of 56
Efficiency / speed	4	7.1%
Verbal fluency / spontaneity	4	7.1%
Accessibility / typing burden	2	3.6%

The minority is not noise; it is a coherent profile. Speed: *"I say what I think and it's done. The app can process it afterwards. Writing makes me spend ten minutes polishing something that should take two"* (P004); *"If I'm on my phone between meetings, it's much easier to say a thoughtful answer than type it with my thumbs"* (P052). Fluency: *"When I type, I start editing myself before I've even finished the thought"* (P017); *"With recording I give the first concrete example that comes to mind instead of turning everything into corporate language"* (P037). Typing burden: *"I type slowly and I over-correct everything. Give me a microphone and let the system turn it into usable feedback"* (P022).

Notably, several audio-choosers explicitly invoked the processing layer in their reasoning: "the app can structure it afterwards," "the system is analysing it rather than sending the recording." For this group the input frame did real work: the pipeline's existence is what makes spontaneous speech acceptable as formal feedback.

5.4 The recipient is almost absent from the reasons

Section 3 set up the discriminating test. The result is one-sided: **none of the 46 writing rationales depended on the assessed colleague hearing the recording**, and respondents knew none would.

The concerns that superficially point at a listener turn out, on inspection, to point elsewhere. Tone concerns (P018, P028, P048) were about what the recording would contain and what the system would extract from it. The anonymity concern (P049) was about the voice identifying its speaker *inside the employer's system*, not to the beneficiary, who never hears it. The performance anxiety (P002, P012, P025) needs no audience beyond the microphone itself.

This is the counterintuitive core of the study. The conversation intuition, together with the reasonable-sounding argument in Section 1 that removing the listener should remove the reluctance, mispredicts the outcome, because it locates the cost of speaking in the listening. For these respondents the cost sits in the recording: the moment of capture and the artifact that survives it.

One honest caveat. Preferences formed over years of message-mode media (voice notes that humans do hear) plausibly carry over as habit even when the listener is removed. P045's *"voice feels a bit too intimate for an HR process"* reads like exactly such imported reasoning. With a single stated-preference question we cannot separate evaluated preference from imported habit; the behavioral study in Section 10 can.

---

6. Against the literature: confirmations, additions and gaps

A single-question interview study earns its keep by being placed against stronger evidence. This section makes the comparison explicit at three levels: the headline number, the individual themes, and the questions neither literature answers. The summary is easy to state: **nothing in our data contradicts the survey-methodology evidence, and two of our highest-ranked themes do not appear in it at all.**

6.1 The headline number

Open-ended survey questions are already the no-listener condition: nobody "receives" a survey answer as a personal message. That literature therefore provides an independent benchmark for our scenario.

Study	Text share	Setting
This study	82% (46/56)	360 feedback about a colleague; app-processed audio
Revilla & Couper [2026], choice condition	~78–83% by question	Narrative open questions; text vs dictation vs recording
Meitinger et al. [2022]	94%	Probability-based panel; text vs voice recording
Lenzner & Höhne [2022]	majority unwilling to use voice	Stated willingness, smartphone surveys

The correspondence is close. Adding a salient human subject, feedback about a colleague with all its social stakes, does not move the needle away from the survey baseline, which is what one would expect if the governing mechanisms are capture-side ergonomics, self-relation to one's voice, and storage concerns rather than anything about who ultimately reads the output. Among voice answerers in the panel study, 54% said afterwards they would rather have typed [Meitinger et al., 2022, in Revilla & Couper, 2026]: even the choosing minority is soft.

6.2 Theme by theme

The table maps each interview theme onto the closest finding in the literature and gives a verdict. Five themes confirm what survey methodology established with larger samples and behavioral

designs; two are effectively new as adoption drivers in this context; the three audio-side themes confirm the known minority profile.

Interview theme (n)	Closest finding in the literature	Verdict
Editability / control (8)	Rehearsability: some media let the sender fine-tune before transmission [Dennis et al., 2008]; written answers are more consciously produced [Gavras et al., 2022]	Confirms
Reflection / structure (7)	Same rehearsability mechanism; writing supports deliberate composition [Dennis et al., 2008; Gavras et al., 2022]	Confirms
Reviewability / clarity (4)	Reprocessability [Dennis et al., 2008]; visible transcripts gave respondents control that stored audio lacks [Revilla et al., 2020]	Confirms
Privacy / environment (6)	Leading reported reason for avoiding voice input: other people being present [Revilla & Couper, 2024]	Confirms
Habit / accessibility (2)	Preference for written communication; voice appeals mainly to those already comfortable with it [Lenzner & Höhne, 2022]	Confirms
Voice / data privacy (5)	Trust in confidentiality predicts voice uptake [Revilla & Couper, 2024]; voice is identifiable behavioral biometric data [Hanisch et al., 2025]	Confirms; sharpened by the employer context
Psychological safety / self-presentation (7)	Selective self-presentation in editable media [Walther, 2007]; absent from the survey literature's catalogue of reasons for modality choice	New in this context
Voice / recording discomfort (7)	Self-voice sounds unfamiliar for physiological reasons [Orepic et al., 2023]; known as perception, not catalogued as an adoption driver	New in this context
Efficiency / speed (4, audio)	Voice lowers completion effort, especially on mobile [Revilla & Couper, 2026]	Confirms
Verbal fluency / spontaneity (4, audio)	Oral answers are more intuitive, longer, more personal [Gavras et al., 2022; Galasso et al., 2024]	Confirms
Accessibility / typing (2, audio)	Typing burden motivates voice and dictation options [Revilla & Couper, 2026]	Confirms

6.3 What the assessment context adds

The two "new" rows are not anomalies; they are what happens to a known choice when the stakes change, and they are the most interesting thing our small study has to offer the survey literature.

Self-presentation moves up the ranking. Survey respondents describe nursing homes or political attitudes to a research agency; nothing they say can cost them a working relationship, so impression management has no reason to surface in their stated motives. Our respondents were composing an evaluation of a colleague inside their own company, and managing how they come across (measured, professional, not impulsive) tied for the second most common primary theme (7 of 56). Walther's selective self-presentation account [2007] predicts exactly this interaction between identity-relevant content and editable media; to our knowledge no voice-input study has content identity-relevant enough to trigger it.

The recorded self-voice becomes structural, not cosmetic. Adoption studies catalogue context and difficulty of oral expression as the main obstacles [Revilla & Couper, 2024]; aversion to hearing one's own recording barely features. In our data it is a top-three theme (7 of 56), and the stated mechanism is specific: quality-controlling spoken feedback means listening to yourself (P002). In a low-stakes survey you can skip the replay; in an assessment of a colleague, review before submission is the responsible thing to do, so an aversion that is cosmetic in surveys becomes a structural barrier

here. The perception literature explains why the aversion exists [Orepic et al., 2023]; the assessment context explains why it matters.

The data controller is the employer. Confidentiality trust predicts voice uptake in surveys [Revilla & Couper, 2024], but a research agency holding an anonymous clip is a different proposition from your employer holding a recognizable recording of your voice discussing a named colleague (P049: "my voice is basically my name"). Same mechanism, higher voltage.

We state the negative result as plainly as the positive ones: **we found no theme that contradicts the survey-methodology evidence.** The differences are of rank and stakes, not of direction. For a stated-preference study with one rater, agreeing with stronger designs where they overlap is the best available sign that the two new themes are worth taking seriously.

6.4 Gaps neither literature covers

Placing the two bodies of evidence side by side also shows what is missing from both. Each gap below maps onto an item of the agenda in Section 10.

- **Behavioral modality choice in workplace assessment.** The survey studies are behavioral but not about assessment; ours is about assessment but not behavioral. Nobody has logged what employees actually do in a live campaign offering both modalities.
- **Record-then-edit as a distinct condition.** The literature tests live dictation and raw recording; recording followed by an editable transcript, which is the design §8 recommends, is a third condition that has not been isolated.
- **Retention policy as a treatment.** Confidentiality trust is measured as a covariate, never manipulated. Whether a credible "raw audio deleted on transcription" guarantee shifts choice is untested.
- **The no-listener property itself.** No study manipulates whether respondents believe a human will hear them. Survey settings imply it; nobody has made it salient and measured the effect.
- **Language and accent.** The voice-input literature is dominated by single-language panels; modality choice by non-native speakers, for whom our data shows text is a precision tool (P046) and voice an exposure (P025), is unstudied.
- **Information quality end-to-end.** The literature measures the richness of raw answers; no one has measured whether that richness survives a consolidation pipeline and improves the document the beneficiary actually reads.

---

7. The trade-off between adoption and information

A consolidation pipeline wants raw material that is specific, behavioral and unlaundered: what happened, who did what, what the impact was. On the evidence, speech is the better carrier of exactly that. Spoken answers to open questions are longer, more spontaneous and contain more personal experience [Gavras et al., 2022; Galasso et al., 2024]; on sensitive topics, voice answers were longer and richer, at the cost of much higher break-off [Höhne et al., 2024]. Our audio-choosers said the same thing from the inside: speaking produces the first concrete example instead of "corporate language" (P037). In our companion audit of an AI-native 360 pipeline, we argued that over-polished, valence-managed text is a validity problem for feedback instruments [Biancheri, 2026]; spontaneous speech is, plausibly, part of the antidote.

But the modality that maximizes information is the one 82% of respondents avoid, and the survey literature adds that pushing voice on people produces break-off, not compliance [Revilla & Couper, 2026; Höhne et al., 2024]. Force voice and you lose respondents; force text and you lose the

respondents for whom typing is the barrier, and the spontaneity of everyone else. **The design problem is therefore not to pick the winning modality. It is to lower the capture-side costs of the losing one for the minority it serves, without imposing it on anyone.**

The interviews say precisely where those costs are: no revision (19 of 46), the recorded self-voice (7), the room you have to find (6), the artifact the company keeps (5). Every one of these is addressable by design, and none of them requires evangelizing anyone into a modality they distrust.

---

8. Design implications for app-processed feedback

We state these as design principles for any system that accepts spoken feedback as input. Nothing below requires our implementation.

1. **Text is the default; voice is an offer, never a requirement.** Every choice study, including this one, puts voluntary voice uptake at a fifth of users or less. A voice-first or voice-only flow is a break-off generator.
2. **Transcribe visibly, and let the respondent edit the transcript before submission.** This single step converts recording into "dictation with control" and neutralizes the largest cluster of objections, revision, review and precision (19 of 46), by giving speech the rehearsability and reprocessability of text [Dennis et al., 2008]. Visible, editable transcription also gave respondents a sense of control in early voice-input survey work [Revilla et al., 2020].
3. **Treat raw audio as toxic inventory: transcribe, then delete by default, and say so at capture time.** Voice is behavioral biometric data [Hanisch et al., 2025]. "The recording is deleted the moment transcription completes" answers P008, P020, P033, P049 and P050 in one sentence. Retention of raw audio should be an explicit, justified exception, not a default.
4. **State at the point of capture that the assessed person never receives the recording.** The property is counterintuitive; do not assume employees know it, or believe it. Some residual message-frame reluctance may rest on simply not trusting this.
5. **Design for the open office.** Typing is discreet; speaking is not. Voice capture is realistically a private-moment, personal-device affordance: support it there, and never build a flow that assumes people can talk about a colleague at their desk.
6. **Do not confuse richer with better.** Spontaneous speech helps only if the pipeline preserves specificity and valence downstream. A pipeline that softens or launders what the speaker actually said re-introduces at the processing stage the distortion it avoided at capture [Biancheri, 2026].
7. **Instrument modality choice in production.** Uptake, completion, break-off, answer length, specificity and valence by modality, the behavioral data this paper lacks, is cheap to log and settles arguments no interview study can.

---

9. Limitations

We list these plainly, because several are structural.

1. **Stated preference, not behavior.** Respondents chose in a described hypothetical, without touching an interface. The survey literature shows stated attitudes and actual uptake diverge in both directions: 54% of actual voice users in one study said they would rather have typed. Nothing here measures what people do in a live campaign.

2. **Single interviewer, single rater, not blind, no reliability statistic.** The author conducted the interviews, extracted the responses from the recordings, and coded them, knowing the research question. The coding frame is published in Appendix B so that this can be contested or replicated by someone else.
3. **Convenience sample.** 56 professionals from the author's network, skewed toward office and knowledge work; demographics not systematically recorded. The 82% is a property of this sample, not an estimate of a population parameter, and should not be cited as one.
4. **One question, one minute.** Primary-theme coding flattens mixed motives; counts are coarse. A respondent citing editability may also hold privacy concerns that never surfaced.
5. **Paraphrase-level reporting.** Responses were condensed from the recordings by the author; there is no independent transcription and no second listener. Quoted material is faithful in content but not verbatim in wording.
6. **Described, not experienced, options.** Respondents' mental models of "recording into an app" surely varied; some reluctance or enthusiasm may attach to imagined interfaces rather than real ones.
7. **Vendor authorship.** The author builds an AI-native feedback platform that accepts both input modes. We note that the result, a strong preference against the more novel modality, is not the commercially flattering one, and we report it anyway; the reader should nevertheless discount accordingly.

---

10. Open research agenda

Each of these is a study we would value running with an academic partner; several we cannot run alone.

1. **The behavioral study.** A live 360 campaign with randomized modality offering (text-only; voice-only; free choice; choice with transcript-editing), measuring uptake, break-off, answer length, specificity, valence, and downstream report quality. This settles what a stated-preference design can only raise.
2. **Does edit-before-submit close the gap?** The dictation/recording distinction predicts that visible, editable transcription recovers much of text's appeal. Test the preference and the quality effects of that single design change.
3. **Does a credible deletion guarantee shift choice?** Randomize the retention statement ("raw audio deleted on transcription" vs. silence) and measure modality choice. This isolates the storage mechanism (5 of 46 primary; plausibly latent in many more).
4. **Does believing "no one hears it" matter at all?** Manipulate the salience and credibility of the no-listener property. If choice does not move, capture-side mechanisms fully dominate; if it does, the message frame retains force.
5. **Language and accent.** Non-native speakers in our sample used text for precision (P046) while accent self-consciousness pushed the same direction (P025). Whether voice input systematically disadvantages non-native speakers, or advantages them once transcription is good enough, is open and consequential for multinational campaigns.
6. **End-to-end information quality.** Do voice-originated consolidated reports contain more specific, behavioral, actionable content than text-originated ones, and does any gain survive the pipeline's own processing, given the valence-steering failure mode documented in [Biancheri, 2026]?

11. Conclusion

When an app processes feedback, spoken audio stops being a message and becomes an input. The decades-old advice to deliver feedback rich is about messages; it says nothing about recordings no one will hear, and the plausible-sounding inference that removing the listener removes the reluctance turns out to be wrong. Offered the choice, 46 of 56 professionals chose typing, for reasons that live in the capture moment and the stored artifact: control over wording before it leaves, the strangeness of one's own recorded voice, the open office around the microphone, and an identifiable recording sitting in an employer's system. The survey-methodology literature, which tests the same choice with no recipient at all, lands on nearly the same number.

For builders the conclusion is directly actionable: default to text, offer voice, transcribe visibly and let people edit, delete raw audio by default and say so. For researchers, the interesting questions are behavioral, and we have specified them. The result we would most like to see tested is the trade-off itself: the modality people avoid is the one that yields the richest raw material for the machine that is listening, the only listener left.

This is an exploratory study (one question, one rater, stated preferences) and we would rather state that plainly and invite scrutiny than overclaim.

References

- Biancheri, A. (2026). When the Model Writes the Questionnaire: An Item-Level Defect Taxonomy for LLM-Generated 360-Degree Feedback Instruments. Preprint v1.
- Bracken, D. W. & Rose, D. S. (2011). When does 360-degree feedback create behavior change? And how would we know it when it does? *Journal of Business and Psychology*, 26(2), 183–192.
- Daft, R. L. & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554–571.
- Dennis, A. R., Fuller, R. M. & Valacich, J. S. (2008). Media, Tasks, and Communication Processes: A Theory of Media Synchronicity. *MIS Quarterly*, 32(3), 575–600.
- Galasso, V., Nannicini, T. & Nozza, D. (2024). We Need to Talk: Audio Surveys and Information Extraction. *CESifo Working Paper* No. 11530.
- Gavras, K., Höhne, J. K., Blom, A. G. & Schoen, H. (2022). Innovating the collection of open-ended answers: The linguistic and content characteristics of written and oral answers to political attitude questions. *Journal of the Royal Statistical Society: Series A*, 185(3), 872–890. doi:10.1111/rssa.12807
- Hanisch, S., Arias-Cabarcos, P., Parra-Arnau, J. & Strufe, T. (2025). Anonymization Techniques for Behavioral Biometric Data: A Survey. *ACM Computing Surveys*. doi:10.1145/3729418
- Höhne, J. K., Gavras, K. & Claassen, J. (2024). Typing or Speaking? Comparing Text and Voice Answers to Open Questions on Sensitive Topics in Smartphone Surveys. *Social Science Computer Review*. doi:10.1177/08944393231160961
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.

- Kumar, A. & Epley, N. (2021). It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal modes of interaction. *Journal of Experimental Psychology: General*, 150(3).
- Lenzner, T. & Höhne, J. K. (2022). Who is willing to use audio and voice inputs in smartphone surveys, and why? *International Journal of Market Research*. doi:10.1177/14707853221084213
- Meitinger, K., van der Sluis, H. & Schonlau, M. (2022). Implementing voice-recordings in a probability-based panel: what we learnt so far. *CIPHER Conference*; figures as summarized in Revilla & Couper (2026).
- Orepic, P., Kannape, O. A., Faivre, N. & Blanke, O. (2023). Bone conduction facilitates self-other voice discrimination. *Royal Society Open Science*, 10(2), 221561. doi:10.1098/rsos.221561
- Revilla, M. & Couper, M. P. (2024). Exploring respondents' problems and evaluation in a survey proposing voice inputs. *methods, data, analyses*. doi:10.12758/mda.2024.06
- Revilla, M. & Couper, M. P. (2026). Expanding Survey Response Options: Combining Dictation and/or Voice Recording With Text to Answer Narrative Open-Ended Survey Questions. *Survey Research Methods*, 20(1), 19–32. doi:10.18148/srm/2026.v20i1.8456
- Revilla, M., Couper, M. P., Bosch, O. J. & Asensio, M. (2020). Testing the Use of Voice Input in a Smartphone Web Survey. *Social Science Computer Review*. doi:10.1177/0894439318810715
- Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23(5), 2538–2557.

Appendix A. The question as posed

The scenario, standardized across all 56 respondents:

> *"Your company is running a 360-degree feedback campaign, and you are asked to take part as a feedback giver for a colleague you work with. You have two options: write your feedback, or record it as audio. In both cases the app processes your input. The person being assessed receives a consolidated written document at the end; they do not receive or listen to your recording. Which would you choose, and why?"*

Procedure: the two options and the processing clarification were read to every respondent in the same wording; each respondent was given one minute to think before answering; the answer consisted of a choice plus free-form reasons. Interviews were conducted individually and audio-recorded.

Appendix B. Coding frame

Each response was assigned the single primary theme whose criterion best matched the reason the respondent led with or stressed. Themes marked (A) were observed only among audio-choosers.

Theme	Criterion for coding
Editability / control	The reason centers on revising, correcting or fine-tuning the message before submission.
Reflection / structure	Time to think, pause, structure or calibrate the content before it goes.
Reviewability / clarity	Re-reading or scanning the answer, checking it against the question; the reader's ability to scan.

Theme	Criterion for coding
Psychological safety / self-presentation	Managing how one comes across: tone, impulsiveness, professionalism, emotional leakage.
Voice / recording discomfort	Aversion to one's own recorded voice, performance anxiety, accent self-consciousness, disfluencies.
Privacy / environment	The physical setting: being overheard, needing a private room, discretion in shared spaces.
Voice / data privacy	The recording as stored, identifiable, quasi-biometric data in the employer's system.
Habit / accessibility	Familiarity with written workflows; language control for non-native speakers.
Efficiency / speed (A)	Faster capture, lower effort, mobile convenience.
Verbal fluency / spontaneity (A)	Thinking aloud, natural expression, avoiding over-editing during capture.
Accessibility / typing (A)	Typing itself as the burden: speed, fatigue, over-correction.