When Photos Talk: The Surprising Realities of Audio in iPhone Live Photos

Zhao Zhao
School of Computer Science
University of Guelph
Guelph, Ontario, Canada
zzhao24@uoguelph.ca

Abstract

Live Photos blur the traditional boundary between photography and videography by recording short sequences of motion and audio around each still image. Although this feature has been promoted as an immersive way to capture life's fleeting moments, the automatic inclusion of sound in "photos" defies the longstanding assumption that photographs are silent. This paper presents findings from a mixed-method study—surveying 212 iPhone users and interviewing 15 of them—to investigate how people perceive, manage, and sometimes regret the embedded audio in Live Photos. While some participants delight in the brief clips of laughter, ambient music, or nature sounds, the majority deem the audio track unnecessary or risky, especially when it includes private conversations. By examining both everyday practices and notable incidents of accidental disclosure, this study reveals that "more data" is not always better.

CCS Concepts

• **Human-centered computing**; • **Human–computer interaction**; • **Empirical studies in HCI**;

Keywords

Live Photos, Multimedia Privacy, User Perception

ACM Reference Format:

Zhao Zhao. 2025. When Photos Talk: The Surprising Realities of Audio in iPhone Live Photos. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3706599.3720006

1 Introduction

Photography has long been cherished as a silent medium, capturing still moments without the accompaniment of sound. The essence of a photograph lies in its ability to freeze time, preserving a visual snapshot that can be revisited without any auditory context. In contrast, videography inherently involves both motion and sound, creating dynamic narratives that unfold over time. This clear distinction between photos and videos has shaped user expectations and social norms for decades: when someone shares a "photo," it is almost universally assumed to be silent, whereas a "video" implies both visual and auditory engagement.

In 2015, Apple introduced Live Photos with the iPhone 6S, a feature that challenged this entrenched distinction by capturing a short video clip—approximately 1.5 seconds—alongside each still image. Live Photos automatically record motion and ambient sound before and after the shutter is pressed, effectively turning static images into brief, animated clips with audio. This innovation was marketed as a way to enhance memories, offering a more immersive and vivid recollection of moments by preserving not just the visual but also the auditory environment of the captured scene.

However, this blending of photo and video introduces complexities that were not fully anticipated. Users who have traditionally relied on photos as silent representations may find themselves unintentionally sharing background conversations, ambient noises, or other private sounds when they share Live Photos. This unanticipated audio layer can lead to unintended privacy disclosures, social embarrassment, and a general sense of unease among both the sharers and recipients of these "talking photos."

The introduction of Live Photos raises several critical questions:

Perception of Value: To what extent do users see embedded audio in Live Photos as an enhancement, and under what circumstances do they consider it beneficial, trivial, or even detrimental?

Privacy and Awareness: How does automatic audio capture raise concerns about accidental disclosure, and to what extent are users aware of potential risks or ways to mitigate them?

Design and Policy: What interventions—ranging from interface cues to default muting—might reconcile the creative potential of Live Photos with the need for user privacy and control?

This study aims to explore these questions through a mixed-method approach, combining quantitative data from an online survey with qualitative insights from in-depth interviews. By doing so, we seek to understand not only the prevalence of Live Photo usage and user attitudes toward its audio component but also the personal experiences and challenges that arise from this feature's integration into daily life.

The importance of this research lies in its potential to inform the design of future multimedia features. As technology continues to evolve, blending different media forms can offer richer user experiences but also introduce new forms of complexity and risk. Understanding how users navigate these changes is crucial for creating tools that respect user expectations and protect privacy without stifling innovation.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/2025/04
https://doi.org/10.1145/3706599.3720006

2 Related work

2.1 The Silent Tradition of Photography

Photography's enduring allure lies in its ability to capture moments in time, offering a silent and still representation of a scene. From its inception in the 19th century, photography has been distinguished from videography by its static nature. The photograph serves as an immutable record, preserving a single instant without the temporal and auditory dimensions that define video [1]. This distinction has fostered specific social norms and user expectations: photos are curated, often shared in controlled settings, and devoid of sound, making them versatile for various uses—from personal memory-keeping to professional documentation [2].

2.2 Multimedia, Metadata, and Privacy Risks

Human–Computer Interaction (HCI) research has extensively explored how additional data layers—such as geotags, timestamps, or metadata—can enrich user experiences but also introduce complexities [4]. Live Photos' embedded audio can be viewed as a form of "invisible metadata," automatically attached to each image without explicit user consent at the moment of capture [5]. This audio layer, while offering enhanced memory recall, can inadvertently capture sensitive or private conversations, ambient noises, and other unintended sounds [5].

Studies on multimedia sharing have highlighted that additional data can provide context and depth to shared content, making it more engaging and informative [6]. However, when this data is not adequately controlled or disclosed, it can lead to privacy breaches and social discomfort. For instance, geotags can reveal a user's location history, while embedded audio might disclose personal conversations or ambient sounds that were never meant to be shared publicly [6]. The automatic nature of Live Photos' audio capture exacerbates these risks, as users may remain unaware that their "photos" contain sound, leading to accidental disclosures when these images are shared [7].

2.3 Context Collapse and Self-Presentation

The concept of context collapse—where diverse audiences converge in a single social media space—has been instrumental in understanding how individuals manage self-presentation online [8]. Live Photos intensify this phenomenon by merging visual and auditory data, thereby complicating the boundaries of what is appropriate to share across different social contexts [8]. For example, a user might intend to share a serene landscape photo with friends but unknowingly include background conversations or noises that reveal more personal or sensitive information [8].

Erving Goffman's framework on the presentation of self-posits that individuals curate their self-image based on the audience and context [9]. Live Photos disrupt these curated presentations by embedding unintentional audio elements that might not align with the intended portrayal. This misalignment can lead to social friction, embarrassment, and privacy concerns, especially when the audio reveals information that was never meant for public consumption [8].

2.4 Tensions Between "More Data" and Usability

Norman's principle of "less is more" in design emphasizes the importance of simplicity and user control [10]. In the context of Live Photos, the addition of audio data aims to enrich the user experience by providing a fuller sensory capture of moments. However, this abundance of data can backfire if users perceive it as intrusive or irrelevant [10]. The challenge lies in balancing the benefits of enhanced memory recall with the potential risks of unintended data exposure [11].

Prior research in HCI underscores that features which automatically add data layers without explicit user consent can lead to decreased trust and increased privacy concerns [12]. Users value control over what information is shared and how it is presented, and when this control is diminished, it can result in negative perceptions of the feature [12]. Live Photos exemplify this tension, as the automatic capture of audio adds complexity to the user experience without providing adequate means for users to manage or opt out of this feature [13].

2.5 Research Gap and Study Contribution

Despite the growing interest in multimedia sharing and user privacy, there remains a notable gap in empirical research specifically addressing user perceptions and experiences with audio in hybrid media formats like Live Photos. Existing studies have predominantly focused on metadata layers such as geolocation or timestamping, leaving the unique challenges posed by embedded audio less explored. Furthermore, the interplay between user expectations of silent photography and the reality of audio-enabled images has not been thoroughly examined. This research seeks to fill these gaps by systematically investigating how users perceive, manage, and sometimes regret the presence of audio in Live Photos. Through a comprehensive mixed-method approach, we aim to uncover the nuanced ways in which this feature impacts user privacy, social interactions, and overall satisfaction with their photographic experience. By doing so, we provide actionable insights and design recommendations that can inform future developments in multimedia features, ensuring they align with user needs and privacy expectations.

3 Methodology

3.1 Study Design and Approach

This research employed a mixed-methods approach, integrating both quantitative and qualitative methodologies to comprehensively explore user perceptions and experiences with audio in iPhone Live Photos. The sequential explanatory design was chosen to first capture broad quantitative trends through an online survey, followed by in-depth qualitative insights from semi-structured interviews. This approach allows for the identification of general patterns and the subsequent exploration of underlying reasons and contextual factors influencing these patterns [1]. The study spanned three months, ensuring sufficient time for data collection, analysis, and iterative refinement of research instruments based on preliminary findings.

3.2 Ethics and Participant Recruitment

All research activities were conducted in strict adherence to ethical guidelines established by the Research Ethics Board (REB) of the University of Guelph, with the study receiving full approval. Participants were provided with detailed information about the study's objectives, procedures, potential risks, and benefits through an informed consent form, ensuring transparency and voluntary participation. Consent was obtained electronically before participants could proceed with the survey or interview.

Recruitment strategies were multifaceted to ensure a diverse and representative sample of iPhone users. The study leveraged various channels, including social media platforms (e.g., Facebook, Twitter, Instagram), university mailing lists, and professional networks, to disseminate survey invitations. Additionally, a snowball sampling technique was employed. To incentivize participation and enhance response rates, participants were offered a chance to enter a raffle for gift cards and received small monetary compensations for completing interviews.

3.3 Phase 1: Online Survey

3.3.1 Survey Development and Instrumentation. The online survey was meticulously crafted to address the study's primary research questions, drawing on established theories and prior research in multimedia sharing, privacy dynamics, and smartphone feature adoption [2]. The survey instrument comprised 35 questions segmented into four main sections: demographics, Live Photo awareness and usage, attitudes toward audio, and privacy concerns and behaviors. Details in appendix.

3.3.2 Distribution and Sampling. The survey was deployed using a reputable online survey platform (Qualtrics), ensuring accessibility across various devices, and maintaining data integrity. Over a four-week period, the survey link was actively promoted through targeted advertisements on social media, emails to university communities, and posts in technology-focused online forums. A total of 248 responses were initially collected. After applying exclusion criteria to remove incomplete submissions and responses from non-eligible participants, 212 valid and complete responses were retained for analysis, yielding an 85% completion rate.

3.3.3 Data Analysis. Quantitative data were analyzed using IBM SPSS Statistics software. Descriptive statistics (means, medians, frequencies, standard deviations) were computed to summarize awareness levels, usage patterns, perceived audio value, and privacy concerns. Pearson's correlation coefficients were calculated to explore relationships between perceived audio enhancement and privacy concerns, while Chi-square tests were conducted to examine differences in audio perception across demographic groups. Additionally, regression analyses were performed to identify predictors of privacy concerns based on usage patterns and attitudes. Open-ended responses were subjected to inductive thematic analysis to identify recurring themes and narratives. Two researchers independently coded a subset of 98 responses, initially identifying codes related to positive audio experiences (e.g., emotional enrichment) and negative experiences (e.g., accidental disclosures). Discrepancies between coders were resolved through discussion and consensus, leading to the development of a robust coding framework that captured the nuanced sentiments expressed by participants.

3.4 Phase 2: In-Depth Interviews

3.4.1 Interview Design and Objectives. Building on the survey findings, phase two involved conducting semi-structured interviews to gain deeper insights into user experiences with Live Photo audio. The interviews aimed to understand the contexts in which audio was perceived as valuable or problematic, understand the emotional and social impacts of accidental audio disclosures, and gather user-driven suggestions for design improvements. This qualitative phase was essential for uncovering the underlying reasons behind the quantitative patterns observed in the survey data.

3.4.2 Participant Selection. From the pool of 212 survey respondents, 40 participants were initially identified for potential interviews based on their varied levels of awareness, usage frequencies, and attitudes toward Live Photo audio. Purposeful sampling was employed to ensure representation across different demographic segments, usage intensities, and privacy concern levels. Ultimately, 15 individuals consented to participate, encompassing a diverse range of perspectives—from enthusiastic adopters who frequently used Live Photos to cautious users who rarely engaged with the audio feature.

3.4.3 Interview Procedure. Each interview was conducted via Zoom, adhering to the IRB-approved protocol to maintain participant confidentiality and comfort. Interviews lasted between 45 to 60 minutes and were audio-recorded with participant consent for accurate transcription. The interview guide was designed to be flexible, allowing for probing and follow-up questions based on participants' responses. Key topics included:

Initial Experiences: Participants were asked to describe their first encounters with Live Photos and how they discovered the audio feature.

Memorable Incidents: Interviewees shared specific instances where the audio in Live Photos either enhanced their experience or led to unintended consequences.

Privacy Concerns: Discussions focused on any feelings of embarrassment, privacy breaches, or social friction resulting from audio disclosures in shared photos.

Design Feedback: Participants provided suggestions for improving the Live Photos feature, emphasizing desired controls, indicators, and user interface enhancements to better manage audio.

3.4.4 Data Analysis. Transcripts from the interviews were meticulously transcribed verbatim and analyzed using Braun and Clarke's (2006) thematic analysis framework [3]. Inter-rater reliability was maintained through regular calibration meetings and consensus discussions, ensuring that the thematic framework was both comprehensive and reflective of the participants' perspectives.

4 Findings

4.1 Survey Results

4.1.1 Awareness and Usage Patterns. The survey revealed that a substantial majority of participants (71%) were aware that Live

Photos included audio. However, only 46% of these users had intentionally explored or utilized this feature beyond casual use. A significant portion (52%) reported sharing at least one Live Photo within the past month, primarily through private messaging apps or personal social media accounts. Daily sharing was relatively uncommon, with only 8% engaging in daily Live Photo sharing.

4.1.2 Perceived Audio Value. When asked to rate the statement "The audio in Live Photos enhances my experience" on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), the average response was 2.4 (SD = 1.1). Approximately 61% of respondents selected 1 or 2, indicating minimal to no perceived benefit from the audio track. Conversely, 17% rated it positively (4 or 5), citing emotional enrichment and enhanced memory recall as key benefits. The remaining participants were neutral or expressed ambivalence. Open-ended responses further elucidated these findings. Users who valued the audio often mentioned specific scenarios, such as capturing a child's laugh or the sounds of nature, which added a layer of authenticity to their memories. In contrast, those who found the audio redundant or intrusive described it as capturing trivial noises like background chatter, footsteps, or the photographer's own prompts.

4.1.3 Privacy Concerns. Over half of the survey respondents (54%) expressed moderate to high concerns regarding the inadvertent sharing of private or sensitive audio content. Among these, 22% reported having experienced at least one incident where the audio in a shared Live Photo led to embarrassment or unintended disclosure. Common examples included overheard conversations, unintentional recording of negative remarks about others, or capturing personal discussions that were not meant for public sharing. The correlation analysis revealed a mild but significant relationship ($r = 0.26$, $p < 0.05$) between the perceived value of audio and privacy concerns. Users who found the audio more beneficial were slightly more likely to worry about accidental disclosures, suggesting a complex relationship between appreciation for the feature and awareness of its risks.

4.1.4 Muting and User Control. Only 38% of participants reported regularly muting the audio in Live Photos before sharing them, while 26% were unsure how to disable or edit the audio track. The remaining 36% did not know how to mute audio or did not believe it was possible without using third-party tools. This indicates a significant gap in user knowledge and highlights the need for more intuitive controls and clearer guidance within the iPhone's interface.

Participants frequently mentioned that the process of muting audio was not obvious or user-friendly. Many expressed frustrations over having to navigate through menus or use additional apps to manage the audio, detracting from the seamless experience that Apple typically promotes.

4.2 Interview Themes

Theme A: Audio as an Afterthought All interviewees acknowledged that they rarely considered the audio component when taking or sharing Live Photos. This afterthought nature often led to unintentional audio captures. One participant stated, "I didn't even realize my phone was recording sound until I listened to a Live Photo and heard background conversations." Another echoed this sentiment, noting, "Most of the time, I forget the audio is there unless someone points it out."

Theme B: Occasional Moments of DelightA minority of participants (5 out of 15) shared positive experiences where the audio enhanced their memories. Examples included capturing a child's laugh, the gentle rustling of leaves, or spontaneous musical moments among friends. One parent described, "Hearing my baby say her first word in a Live Photo is priceless. It adds so much emotion to the picture." These instances illustrate that audio can provide meaningful context and emotional depth in specific, cherished moments.

Theme C: Frequent Redundancy or Irrelevance Despite those occasional delights, most interviewees found the sound in Live Photos as random background noise—footsteps, television chatter, or the photographer's own instructions ("Say cheese!"). One participant remarked, "I love the movement, but the sound is just random chatter. It doesn't add anything meaningful." This redundancy often overshadowed the perceived benefits of the audio, leading to a general disinterest in utilizing this feature.

Theme D: Accidental Exposure and Privacy Risks Several participants recounted incidents where unintended audio captures led to social embarrassment or privacy breaches. For instance, one user shared, "I took a Live Photo at a family dinner, and someone was discussing their job in the background. It was awkward when everyone heard it." Another described capturing a friend's offhand negative comments about a mutual acquaintance, which caused tension when the audio was played back in a group chat. These stories highlight the potential for Live Photos to inadvertently reveal sensitive or private information, thereby undermining the user's intent to share innocuous moments.

Theme E: Desire for Clearer Defaults and IndicatorsA unanimous sentiment among interviewees was the need for more transparent controls and indicators regarding audio in Live Photos. Suggestions included Default Muting, Visual Cues, Simplified Controls, and Post-Capture Prompts. These design recommendations aim to empower users with greater control over their shared content, reducing the likelihood of accidental disclosures and enhancing overall user satisfaction with the feature.

5 Discussion

Live Photos represent a significant shift in how users interact with their photographic devices by merging the traditionally silent medium of photography with the dynamic, auditory elements of videography. This convergence disrupts deeply ingrained user expectations, wherein a "photo" is synonymous with a static, soundless image. The study's findings underscore that while the feature introduces novel possibilities for memory preservation, it also introduces substantial risks related to privacy and social norms.

Addressing the dual imperatives of enhancing user creativity and safeguarding privacy requires thoughtful design interventions. The study's findings suggest several key strategies:

- Default Muting: Implementing audio-off as the default setting can prevent accidental disclosures, aligning with

privacy-by-design principles [14]. Users who wish to include audio can opt in, ensuring that the feature respects their privacy preferences.

- Clear Visual Indicators: Introducing visible cues, such as speaker icons or audio watermarks, can remind users and recipients that a Live Photo contains sound. These indicators serve as constant reminders to manage and be mindful of the audio content.
- Simplified Controls: Providing intuitive, one-tap mute options within the Photos app can empower users to manage audio tracks effortlessly. Simplified interfaces reduce the cognitive load and make it easier for users to control their shared content.
- Post-Capture Prompts: Introducing prompts immediately after taking a Live Photo to confirm whether to keep or discard the audio can enhance user awareness and decision-making. These prompts act as a checkpoint, ensuring that users consciously choose to include or exclude audio based on the context of the moment.

These design recommendations aim to align Live Photos more closely with user expectations and privacy norms, reducing the likelihood of unintended disclosures while preserving the feature's creative potential.

The lessons learned from Live Photos extend beyond this specific feature, offering broader insights into the design of multimedia tools. As technology continues to integrate various data layers—visual, auditory, geolocation, and beyond—designers must prioritize user control, transparency, and simplicity to ensure that these features enhance rather than complicate user experiences.

Finally, while this study provides valuable insights into user perceptions of Live Photo audio, we acknowledge it is not without limitations. The sample was predominantly drawn from North American iPhone users, potentially limiting the generalizability of the findings to other cultural or regional contexts where social norms around photography and privacy may differ. Additionally, the self-reported nature of survey and interview data may introduce biases related to memory recall and social desirability.

6 Conclusion

The integration of audio into Live Photos represents a pivotal moment in the evolution of smartphone photography, challenging users' traditional perceptions of what constitutes a photograph. While the feature offers the potential for richer, more immersive memories, it simultaneously introduces significant risks related to privacy and social norms. Our mixed-method study, encompassing both quantitative and qualitative data, reveals that although a subset of users cherishes the added auditory dimension in specific, meaningful contexts, the majority find the audio component redundant or intrusive. Incidents of accidental audio disclosures highlight the vulnerability users face when sharing what they believe to be silent images. These findings underscore the necessity for thoughtful design interventions that prioritize user control and privacy without stifling the creative possibilities that features like Live Photos aim to offer. By implementing default muting, clear visual indicators, and intuitive audio management tools, designers can better align Live Photos with user expectations, ensuring that "more data" enhances rather than complicates the user experience.

References

[1] Take and edit Live Photos - Apple Support (CA). (n.d.). Retrieved from https://support.apple.com/en-ca/104966

[2] Khoshsirat, S., & Kambhamettu, C. (2023). Empowering visually impaired individuals: A novel use of apple live photos and android motion photos. *arXiv preprint arXiv:2309.08022*.

[3] Deogracias, A. (2015). Danah Boyd: It's Complicated: The Social Lives of Networked Teens: Yale University Press, New Haven, Connecticut, 2014, pp. 296, ISBN 973-0-300-16631-6.

[4] Sites, S. N. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications.

[5] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

[6] Ray, R. (2007). Designing and conducting mixed methods research. *Qualitative Research Journal*, *7*(2), 90-92.

[7] Goffman, E. (2023). The presentation of self in everyday life. In *Social theory re-wired* (pp. 450-459). Routledge.

[8] Goffman, E., *The Presentation of Self in Everyday Life.*, Doubleday, 1959.

[9] Barclay, K., McMahon, D., & Stearns, P. N. (Eds.). (2024). *The Routledge History of Happiness*. Taylor & Francis.

[10] Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

[11] Palen, L., & Dourish, P. (2003, April). Unpacking" privacy" for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 129-136).

[12] McDonald, N., & Forte, A. (2021). Powerful privacy norms in social network discourse. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1-27.

[13] Leonardi, P. M. (2011). Innovation blindness: Culture, frames, and cross-boundary problem construction in the development of new technology concepts. *Organization Science*, *22*(2), 347-369.

[14] Langheinrich, M. (2001, September). Privacy by design—principles of privacy-aware ubiquitous systems. In *International conference on ubiquitous computing* (pp. 273-291). Berlin, Heidelberg: Springer Berlin Heidelberg..