Policy Brief

September 2025

Sound Money

Sensory Infrastructure for Inclusive Digital Payments

Mac Milin Kiran
Cherry Wu

The Future of Digital Finance

Emerging opportunities in India, in China and on the African continent

Centre for International Governance Innovation

T20 SOUTH AFRICA 2025

Key Points

- Confirmation cues — audio, haptic, spoken and visual — are core infrastructure; fragmentation across apps and borders makes “your money moved” non-portable, undermining trust and excluding first-time digital users in cash-intensive economies.
- Evidence from India, China and Kenya shows single-modality cues are fragile and social fit matters; systems need redundant, recognizable signals that travel across devices, languages and settings.
- Authorities should adopt a minimal, open Multisensory Trust Stack, require redundancy proven in real-world conditions and constrain voice with privacy by design and anti-spoofing safeguards.
- Adoption should be scaled through procurement-led specifications and lightweight conformance tests, ensuring confirmation performance is measurable domestically and cross-border.

Introduction

As digital payments become rapidly adopted in retail, transport and social protection, hundreds of millions across the Global Majority become first-time users of formal finance. Policy has focused on fees, Know Your Customer (KYC) and cybersecurity. Yet one element remains underexplored: how people know a payment has gone through. For users in cash-intensive economies, trust is sensory: a beep, a vibration, a green tick, a spoken receipt. Rather than ornamental, these signals are the sensory infrastructure of digital finance.

When that layer is missing, inconsistent or inaccessible, uncertainty rises. People repeat actions, low-vision and low-literacy users are excluded, and fragmented cues across apps and jurisdictions create new interoperability risks (Sun et al. 2010; Pataca et al. 2025).

This brief examines three cases (India, China and Kenya) to show how confirmation cues succeed or fail in context. We then distil cross-case lessons and offer recommendations grounded in those lessons for central banks, standards bodies and regional operators.

Context

Clear, immediate confirmation cues reduce disputes and lighten cognitive load at the point of sale (Kohrs et al. 2016; Sena et al. 2025). When cues are audible or haptic, they support participation by low-vision and low-literacy users and curb the repetitive actions and frictions that can compound into exclusion (Sun et al. 2010). Human–computer interaction research links timely, unambiguous feedback to perceived control and satisfaction; ambiguous or delayed cues drive retry behaviours and elevate support costs (Kohrs et al. 2016; Sena et al. 2025).

As payment systems consolidate domestically and link across borders, confirmation architectures remain fragmented. Users encounter inconsistent tones, phrasing and haptic patterns, and habits formed in one app or device do not transfer to another. This fragmentation particularly affects users transitioning from cash-based systems who rely more heavily on sensory cues for transaction confidence. This is an interoperability gap as real as a messaging mismatch: the final mile of “your money moved” is not yet portable.

Foundations for multimodal, perceivable feedback already exist in adjacent accessibility and payments frameworks, though few explicitly address confirmation cues. For example, the European Accessibility Act requires product interfaces to use more than one sensory channel: when information is displayed visually, an audio or tactile alternative must also be provided (European Accessibility Act 2019). As efforts to harmonise data layers progress, the confirmation layer can be better conceived as complementary infrastructure, meriting more systematic attention rather than ad hoc solutions.

As cross-border systems scale and domestic rollouts expand, the absence of consistent confirmation cues poses a growing risk to trust and inclusion. This context motivates a case-led examination. India, China and Kenya provide distinct angles on the same question: how can confirmation cues be designed so that they work with social practices, environments and diverse abilities? They must serve users with varying literacy levels and digital experience, and travel across providers and borders without confusion.

India: Soundboxes and the Audibilization of Trust

In India, the Unified Payments Interface (UPI) made QR payments ubiquitous for person-to-merchant transactions (National Payments Corporation of India). Yet confirmations were easy to miss in crowded market stalls. A discreet buzz on a customer’s phone or a small success icon on a screen could go unnoticed by the seller. Providers responded with soundboxes: compact speakers that announce each successful transaction in local languages.

Adoption surged. In early 2024, more than 20 million merchants across providers had adopted soundboxes (Singh 2024). Design features evolved with the diversity of India’s retail environment: multilanguage support, louder or crisper audio for noisy markets and, in some models, the integration of tap-to-pay alongside voice alerts. Paytm’s Card Soundbox added NFC with spoken confirmations in eleven languages; PhonePe’s SmartSpeaker offered 4G connectivity and twenty-one languages; Google’s SoundPod was rolled out nationally after a pilot with regional language support (Paytm 2025; PhonePe 2025; Singh 2024).

The trajectory illustrates two points. First, moving confirmation from the customer’s device to a shared, audible cue in the merchant’s space reduced ambiguity, shortened back-and-forth checks and lowered disputes. The message is addressed to both parties at once. Second, audibility is context dependent. The very environments that make a public cue useful —

busy, noisy shops — can also make it hard to hear. A purely audio solution is, by definition, fragile for users with hearing loss and for times when markets are at their loudest.

The deeper lesson from India is social. Soundboxes work because they fit existing merchant workflows. The cue arrives at precisely the moment a seller would otherwise glance at a phone or ask for a screenshot. Over time, the cue becomes an expectation: a local norm that says “the money is in.” That norm, however, is tied to specific devices and provider ecosystems. When a seller switches to a new provider or a buyer pays with a cross-border wallet, tones and spoken phrasing can differ. What scaled trust domestically is not yet portable across providers or borders.

China: Super-Apps, Spoken Receipts and Voice-Forward Accessibility

China’s retail payment ecosystem is dominated by Alipay and WeChat Pay, which process billions of transactions per day (Elad 2025). Decisions by these super-apps often become de facto national standards. For merchants, both platforms provide spoken receipts that announce the amount received (for example, “WeChat Pay receiving 20 yuan”). Research with senior street vendors found the cue to be “the most direct and convenient” way to check payments while continuing to serve customers (He et al. 2023). The feature’s limitations are instructive: in very noisy markets, the message is not always audible; for users with hearing loss, it is unreliable; and when transactions overlap, an amount-only message without payer details can cause confusion.

For consumers, Alipay has layered accessibility features such as voice input for entering amounts, audio playback of the amount to be sent and voiceprint authentication as an alternative to PINs (Business Wire 2023). The timing of upgrades, highlighted during a major event like the Asian Para Games, signals a design stance: audio confirmations are considered core infrastructure for inclusion rather than optional extras.

Voice has also been put to work as a fraud awareness safeguard. In 2025, Alipay introduced an in-app voice call function that displays a counterparty’s verified real name before transfers (Feng 2025). The goal is to combat impersonation fraud by shifting reassurance into the moment of risk rather than relying on post hoc confirmations. At the device frontier, Alipay piloted AR glasses payments with partners such as Rokid and Meizu (Borak 2025). Users scan a QR through the glasses and confirm by voice; the trust signal combines voiceprint authentication with on display prompts, and the company has framed the pilots as a path to extending confirmation cues into wearables and, eventually, cross-border acceptance via Alipay+.

China’s case shows that audio and voice are now mainstream and multifaceted: they support merchant operations, improve accessibility, harden high-risk flows and enable new devices. Yet, as with India, implementations vary by app and device. Spoken phrasing, audio signatures and visual banners are not standardised across the ecosystem. Cues remain non-portable: a recognisable success sound in one app may mean nothing in another, and cross-border users encounter unfamiliar patterns.

Kenya: Voice-Enabled Assistance in Small Business Payments

Kenya remains globally recognised for pioneering mobile money through M-Pesa, where everyday confirmations rely on USSD menus and SMS receipts (Fengler 2012). These channels are robust on basic phones and work in low-connectivity contexts, but they are text heavy and provide only high-level details. For small merchants, reconciling what was actually sold against payment notifications is labour intensive.

A research prototype explored whether voice could lower barriers. The Dukawalla pilot, a collaboration between Microsoft Research Africa, UC Irvine and Kenyan partners, trialled a mobile assistant with seven Nairobi businesses (Ankrah et al. 2025a). Merchants recorded sales and payments by speaking naturally; speech recognition and generative AI turned recordings into structured transaction data and provided spoken summaries.

The results were instructive. Operationally, merchants hoped to save time and reduce manual record keeping, but noisy markets degraded recordings. Merchants struggled to use the tool while serving customers at speed. Culturally, speaking transactions aloud sometimes conflicted with Kenyan business norms built on socio-tecture: the blending of social and commercial life (Ankrah et al. 2025b). Customers questioned why purchases were being verbalised, and some merchants reported that voice entry felt "weird" or intrusive. Technically, Kenya's multilingual reality proved challenging. Merchants code switched between Kiswahili, English and local dialects; colloquialisms such as "bob" (for a shilling) or "one fifty" (meaning 150 shillings) confused recognition models.

The Kenya case underlines both promise and fragility. Voice can reduce literacy barriers and help when screens are small or unreadable, but confirmations must fit social practice, work in noisy, multilingual contexts and avoid disrupting the relational norms of small-business transactions. A design that compels merchants to speak every purchase aloud may clash with privacy expectations and create friction precisely where speed and discretion matter.

Lessons for Policy Makers, Standards Bodies and Operators

What can other countries learn from India, China and Kenya? This brief does not assert that any one country's approach represents the optimal strategy. However, drawing on the insights described above, it recommends that payment authorities, standards bodies, regional operators and implementers adopt a Multisensory Trust Stack as the foundational framework for inclusive payment confirmation.

In practice, this means central banks such as the Reserve Bank of India (RBI), the People's Bank of China (PBoC) and the Central Bank of Kenya (CBK). It includes standards bodies like the International Organization for Standardization (ISO) and EMVCo, with the World Wide Web Consortium (W3C) providing accessibility guidance. It also extends to regional operators, including the National Payments Corporation of India's Unified Payments Interface (NPCI/UPI), the Pan-African Payment and Settlement System (PAPSS) led by the African Export-Import Bank, and the mBridge consortium piloting cross-border settlement.

To address the fragmentation and fragility highlighted in our cases, the Multisensory Trust Stack can be advanced through the following coordinated actions:

Establish the Trust Stack as standardised infrastructure: Fragmented confirmations erode user confidence, particularly for low-literacy and low-vision users who rely on consistent cues to build trust across providers and borders. The Stack should comprise:

- auditory confirmation through standardised spoken receipts and tones;
- haptic fallback such as vibration patterns or LED cues for noisy environments;
- adaptive language using on-device text-to-speech with local dialect switching; and
- an open API layer enabling wallets and central bank digital currencies (CBDCs) to exchange confirmation metadata in real time.

Central banks such as the RBI and CBK could codify this minimal library of primitives, while ISO and EMVCo would be natural convenors for portability and version control.

Mandate the Trust Stack as infrastructure, not ornamentation: Evidence from India and China shows that confirmation cues determine whether users feel "the money moved," yet they are often treated as proprietary features. Because these signals represent the final mile of transaction trust, regulators should integrate the Trust Stack into payment system specifications alongside messaging and settlement protocols. Authorities such as the PBoC and RBI could require compliance for licensing and participation in interoperability schemes, while EMVCo and W3C can provide the technical guardrails to embed accessibility requirements.

Require redundancy and social fit in real-world conditions: Single-modality designs fail predictably across contexts. India's soundboxes struggle in noisy markets; China's visual interfaces exclude low-vision users; and Kenya's text-heavy SMS receipts disadvantage low-literacy users, while voice pilots clash with privacy norms. Trust Stack certification should mandate at least two concurrent confirmation modes (for example, audio plus visual, or audio plus haptic), graceful degradation offline and on low-end devices, and field-tested usability across noise bands, older adults, low-vision populations and local languages. National operators such as NPCI/UPI and CBK's QR Standard unit are well placed to lead this field testing.

Implement voice with privacy by design safeguards: Voice confirmations can reassure users at high-risk moments, as China's fraud awareness features show, but they can also enlarge attack surfaces and raise cultural concerns, as in Kenya's Dukawalla pilot. Trust Stack voice specifications should require template-based speech rather than unconstrained generative output, mandate device-local synthesis where feasible, prohibit audio retention and include anti-spoofing measures such as watermarked tones or device-bound pattern checks, with thresholds for accuracy across accents and latency.

Scale adoption through procurement and measurable standards: Voluntary uptake will not suffice. Regional operators should run domestic and cross-border pilots, making confirmation a first-class metric and publishing results on latency, recognisability, language coverage and failure modes (for example, overlapping transactions, noisy retail and low-end devices). Public purchasers should embed Trust Stack compliance in tenders for POS devices, transit validators and disbursement apps, making standardised confirmations the market default. Regional sandboxes such as India Stack and PAPSS can test interface standards in realistic conditions, while targeted R&D can drive down costs for low-power audio and haptic modules.

Implementing a Multisensory Trust Stack will require upfront investment. Costs include integrating audio or haptic modules into low-end devices, upgrading point-of-sale terminals and localising voice technologies across languages and dialects. Merchants may also face subscription costs for devices such as soundboxes, while providers must invest in speech synthesis and anti-spoofing safeguards. Yet these barriers are not insurmountable. Coordinated procurement can amortise hardware costs, as evidenced in India's bulk purchase of soundboxes to accelerate merchant adoption. Foundations also exist in adjacent frameworks. For example, the aforementioned European Accessibility Act requires product interfaces to use more than one sensory channel, showing how multimodal feedback can be mainstreamed through regulation. With targeted R&D and structured pilots, solutions can likewise be adapted to noisy markets, low-connectivity environments and low-end devices.

The benefits would be felt most immediately in cross-border corridors, where settlement is already instant but confirmation cues still diverge. At present, a sender may see only a push notification while the receiver gets a text, with no shared standard. Multi-jurisdictional linkages illustrate how a Multisensory Trust Stack could close this gap. In the India–Africa corridor, NPCI International has partnered with the Bank of Namibia to license and launch UPI, part of a broader strategy to expand real-time payments into African markets and support financial inclusion (Kaaru 2025). Embedding confirmation cues into such deployments would ensure that migrant workers and small businesses encounter consistent signals at both ends of a transfer. On the China–Africa front, the African Export-Import Bank and the Export–Import Bank of China have established ongoing cooperation to deepen financial integration and trade finance between the two regions (African Export-Import Bank

2023). Such infrastructure provides a natural testbed for portable confirmation cues in multi-jurisdictional flows.

Together, these developments highlight the larger principle: by embedding confirmation cues into the same rails as value, a Trust Stack would make confidence interoperable across jurisdictions, tying the recommendations together as a coherent path forward.

Conclusion

Genuine inclusion in digital payments depends on what people can perceive, particularly first-time users transitioning from cash-intensive societies. The cases of India, China and Kenya show that confirmation cues are not cosmetic. They are the final mile of trust that determines whether users feel confident enough to pay again tomorrow. India demonstrates how a shared audible cue can scale into infrastructure at the point of sale. China shows that voice and audio can anchor accessibility and risk-aware flows but remain non-portable across apps and devices. Kenya cautions that designs which ignore noise, multilingual reality and social practice will falter. The common thread is clear: single-modality cues are fragile, and fragmented cues do not travel.

A practical path is within reach. Payment authorities can specify a minimal, open Multisensory Trust Stack that any wallet, terminal or app can implement. They can require redundancy proven in real-world conditions and constrain voice with privacy by design and anti-spoofing safeguards. Regional operators can make confirmation performance measurable at home and across borders, while public procurement and lightweight conformance tests convert principles into shipping products. Taken together, these steps make "your money moved" portable, recognisable and inclusive. They strengthen trust for low-vision and low-literacy users, support merchants in noisy markets and ensure that the next generation of cross-border rails carries confidence as well as value, embedding digital financial inclusion as core infrastructure for the Global Majority and beyond.

About the Authors

Mac Milin Kiran is a technology and public policy specialist with a background in computer science from Nanyang Technological University. He holds a master's degree in Communication, Culture and Technology from Georgetown University. He has worked on digital governance issues in roles at Georgetown University and with the office of an Indian Member of Parliament. His work has been published by *Tech Policy Press*, the Center for Democracy & Technology, *ProMarket* and the Beeck Center for Social Impact and Innovation. He is currently an Innovation Scholar at the U.S. Chamber of Commerce Foundation. His writing is undertaken in a personal capacity.

Cherry Wu is a technology and public policy specialist with a background in international affairs and AI governance. She holds a master's degree from Georgetown University's School of Foreign Service and a bachelor's degree from McGill University. She has worked on AI and digital policy issues in roles at the Center for Security and Emerging Technology, Lawrence Livermore National Laboratory and the World Bank. She currently leads strategic AI projects at Duco, a technology policy consultancy. Her writing is undertaken in a personal capacity.

Works Cited

"Alipay Enhances Accessible Payment Service in Support of Asian Para Games." *Business Wire*, October 19, 2023. Accessed September 12, 2025. https://www.businesswire.com/news/home/20231018710387/en/Alipay-Enhances-Accessible-Payment-Service-in-Support-of-Asian-Para-Games.

Ankrah, Elizabeth A., Kagonya Awori, Stephanie Nyairo, Mercy Muchai, Millicent Ochieng, Mark Kariuki, Gillian R. Hayes and Jacki O'Neill. "Social by Nature: How Socio-Tecture Shapes the Work of SMBs and Considerations for Reimagining Collaborative Human-AI Systems." *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, April 25, 2025b, 1–19. https://doi.org/10.1145/3706598.3715019.

Ankrah, Elizabeth, Stephanie Nyairo, Mercy Muchai, Kagonya Awori, Millicent Ochieng, Mark Kariuki and Jacki O'Neill. "Dukawalla: Voice Interfaces for Small Businesses in Africa." *arXiv preprint*, May 8, 2025a. https://doi.org/10.48550/arXiv.2505.05170.

Borak, Masha. "Alipay Introduces Smart Glasses Payment with Voice Authentication." *Biometric Update*, June 23, 2025. Accessed September 12, 2025. https://www.biometricupdate.com/202506/alipay-introduces-smart-glasses-payment-with-voice-authentication.

Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the Accessibility Requirements for Products and Services [2019] OJ L 151/70.

Elad, Barry. "Alipay vs. WeChat Pay Statistics 2025: Market Share, Innovation & Digital Yuan Impact." *CoinLaw*, August 3, 2025. Accessed September 12, 2025. https://coinlaw.io/alipay-vs-wechat-pay-statistics/.

Feng, Coco. "Alipay Launches Voice Call to Highlight Financial Security via Real Identity Checks." *South China Morning Post*, May 12, 2025. Accessed September 12, 2025. https://www.scmp.com/tech/tech-trends/article/3309999/alipay-launches-voice-call-highlight-financial-security-real-identity-checks.

Fengler, Wolfgang. “How Kenya Became a World Leader for Mobile Money.” *World Bank Blogs*, July 16, 2012. Accessed September 12, 2025. https://blogs.worldbank.org/en/africacan/how-kenya-became-a-world-leader-for-mobile-money.

He, Changyang, Lu He, Zhicong Lu and Bo Li. “‘I Have to Use My Son’s QR Code to Run the Business’: Unpacking Senior Street Vendors’ Challenges in Mobile Money Collection in China.” *Proceedings of the ACM on Human-Computer Interaction* 7, no. CSCW1 (April 14, 2023): 1–28. https://doi.org/10.1145/3579493.

Kaaru, Steve. “Namibia to Launch a New Payments System Based on India’s UPI.” *CoinGeek*, July 25, 2025. Accessed September 12, 2025. https://coingeek.com/namibia-to-launch-a-new-payments-system-based-on-india-upi/.

Kohrs, Christin, Nicole Angenstein and André Brechmann. “Delays in Human-Computer Interaction and Their Effects on Brain Activity.” *PLOS ONE* 11, no. 1 (January 8, 2016). https://doi.org/10.1371/journal.pone.0146250.

Lacerda Pataca, Caluã de, Saad Hassan, Lloyd May, Michelle M. Olson, Toni D’Aurio, Roshan L. Peiris and Matt Huenerfauth. “Tactile Emotions: Multimodal Affective Captioning with Haptics Improves Narrative Engagement for d/Deaf and Hard-of-Hearing Viewers.” *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, April 25, 2025, 1–17. https://doi.org/10.1145/3706598.3713304.

“Paytm Launches India’s First NFC Card SoundboxTM, a Two-in-One Mobile QR Payment Device That Doubles up as an Affordable Card Payments Machine, for Millions of Offline Merchants.” *Paytm*, May 19, 2025. Accessed September 12, 2025. https://paytm.com/blog/payments/paytm-launches-indias-first-nfc-card-soundbox/.

“PhonePe Unveils Its Made in India SmartSpeaker.” *PhonePe*, May 5, 2025. Accessed September 12, 2025. https://www.phonepe.com/press/phonepe-unveils-its-made-in-india-smartspeaker/.

Sena, Maria Klara, Emilly Yorke Lima and Raul Benites Paradeda. “Timing Matters: Comparing the Effects of Immediate, Delayed, and No Feedback on User Trust in Interactive Systems.” *Anais do XVII Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2025)*, July 20, 2025, 31–40. https://doi.org/10.5753/sbcup.2025.7935.

Singh, Manish. “Google Pay Takes Its QR Soundbox to Small Merchants in India after Trial Run.” *TechCrunch*, February 22, 2024. Accessed September 12,

2025. https://techcrunch.com/2024/02/22/google-pay-takes-its-qr-sound-box-to-small-merchants-in-india-after-trial-run/.

Sun, Minghui, Xiangshi Ren and Xiang Cao. “Effects of Multimodal Error Feedback on Human Performance in Steering Tasks.” *Journal of Information Processing* 18 (2010): 284–92. https://doi.org/10.2197/ipsjjip.18.284.

“UPI: Unified Payments Interface – Instant Mobile Payments: NPCI.” *National Payments Corporation of India (NPCI)*. Accessed September 12, 2025. https://www.npci.org.in/what-we-do/upi/product-overview.