

Who Speaks, Who Receives, and Who Answers? Voice Technology in Crisis Communication

BELU TICONA, George Mason University, USA

XIOMARA HURNI-CRANSTON, CLEAR Global, USA

ANTONIS ANASTASOPOULOS, George Mason University, USA

Speech technologies may bypass literacy barriers in crisis communication, but speech alone does not ensure people can receive, understand, trust, or answer a message. We analyze 35 key informant interviews across ten underserved-language contexts in eight countries. Informants included humanitarian workers, linguists, and technology professionals. Their accounts show that spoken crisis communication depends on five interacting conditions: a speaker whose authority fits the task, a locally meaningful language variety, an available channel, listener control of a device, and organizational capacity to respond. We describe this combination as *relational voice infrastructure*. Across contexts, trusted relationships made communication credible and answerable; orality worked around but did not remove language and literacy inequalities; and voice technology remained constrained by communication practices, infrastructure, and household access. We connect HCI and low-resource speech research and offer a framework for assessing whether voice systems help people understand, trust, question, and act on crisis information.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *Empirical studies in collaborative and social computing*.

Additional Key Words and Phrases: crisis communication; trust; orality; voice technology; underserved languages

ACM Reference Format:

Belu Ticona, Xiomara Hurni-Cranston, and Antonis Anastopoulos. 2027. Who Speaks, Who Receives, and Who Answers? Voice Technology in Crisis Communication. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI '27)*. ACM, New York, NY, USA, 24 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

One practitioner working with Seraiki-speaking communities along the Indus in Pakistan described riverine households with one smartphone, held by a male family member. A flood warning reached everyone else only if he was home, noticed it, and passed it on (L09-NGO-02). Another described remote, disaster-prone communities as both highly exposed and without the electricity, phones, information access, education, and road access that current voice tools assume (L09-NGO-01). A spoken warning removes the need to read. It does not change who holds the handset, whether the power is on, or whether anyone is available when someone calls back.

Crisis-communication research has documented how emergency managers and essential workers use social media and workplace communication to find and circulate information [20, 21], and how displaced communities use WhatsApp voice notes and interactive radio when written or one-way communication falls short [30, 31]. Speech technology extends this possibility by allowing people to speak and listen rather than read and write [14, 23, 25]. Yet a channel's

Authors' Contact Information: Belu Ticona, George Mason University, Fairfax, USA, mticonao@gmu.edu; Xiomara Hurni-Cranston, CLEAR Global, Sandford, USA; Antonis Anastopoulos, George Mason University, Fairfax, USA, antonis@gmu.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2027 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

availability does not by itself make crisis information usable. A warning must be intelligible in the variety people speak, credible because of who is heard saying it, reachable when power and networks fail, and answerable when someone needs clarification or needs to report harm.

HCI and CSCW provide a relational account of these conditions. Crisis information is produced through collective human and technical work [29], while information access and trust are shaped by social relationships, institutions, histories, and technological artifacts [2, 13, 26]. A systematic review of voice-based community engagement similarly identifies moderation, maintenance, and stakeholder authority as conditions of participation alongside system design [14]. Recent work with Rohingya participatory media describes these relations as a *relational media infrastructure*, shaped by material access, gendered risk, and institutional control [10].

Speech research approaches the problem from another direction. Massively multilingual models now report speech-recognition support for more than a thousand languages [19], while Common Voice established a crowdsourced approach to building multilingual speech corpora [3]. Community-centered research nevertheless shows that support for a named language does not establish a system’s fit with local speakers, oral practices, or uses [23, 25]. Prior work therefore identifies many parts of the problem, but provides limited cross-context evidence about how social, linguistic, material, and organizational conditions interact in crisis communication across underserved-language settings.

We take up that question through a qualitative analysis of 35 key informant interviews on crisis communication in ten linguistic contexts across eight countries, conducted with humanitarian practitioners, linguists, and technology-sector informants who work with these language communities. We ask two questions. **RQ1: what conditions do informants describe as shaping whether crisis information reaches, is believed by, and can be answered by the people it is meant for?** **RQ2: which of those conditions recur across linguistic contexts, and which are specific to a setting?**

Across the accounts, whether a spoken message could reach someone depended on five interacting conditions: a speaker whose authority fitted the task, a language and variety the listener used, a channel that survived the crisis, a device the listener controlled, and an organization that had assigned someone to answer. Building on relational accounts of communication infrastructure, we call this arrangement a *relational voice infrastructure*. Accounts from the ten contexts showed different combinations of these conditions failing: a feedback loop could be staffed and still fail on language, while a warning network could be fast, trusted, and locally run and still stop at the people outside it.

We contribute a cross-context empirical account of where spoken crisis communication breaks, and an analytic frame that makes those points of failure visible across HCI and speech research. For speech researchers, the frame shows why language coverage says little about reach without specifying the variety, speakers, task, and conditions of access. For HCI and crisis researchers, it connects communication practices to the linguistic and organizational conditions that determine whether people can understand, question, and act on a message.

2 Related Work

2.1 Trust as a property of arrangements, not tools

HCI research in humanitarian settings treats trust as relational and contextual rather than as an attitude toward a particular technology. Almohamed et al. connect language barriers to refugees’ ability to form relationships in a host community, showing that information access rests on trusted sources, privacy, and social connection [2], and Bautista Isaza et al. describe trust in humanitarian demining as a condition produced through interactions among people, institutions, histories, and information systems [13]. Crisis informatics makes the parallel argument at field level, that

crisis information is produced through social and institutional arrangements rather than delivered through neutral channels [29]. The distinction this work draws, between trusting a tool and trusting the arrangement it sits in, is the one our findings extend to spoken crisis communication.

2.2 Voice as a social arrangement

Work with displaced communities brings that view into language and voice. Talhouk et al. found that Syrian refugee women preferred WhatsApp voice notes to text, with the preference shaped by language and literacy and bounded by connectivity [31], and their study of synchronous community radio documented shifts in trust and agency between refugees and healthcare providers alongside the power and coordination that community-hosted broadcasting involves [30]. Islam et al. synthesize this strand, finding across 29 studies of voice-based community engagement that moderation, maintenance, and stakeholder authority shape outcomes as much as system architecture does [14]. Voice matters here not as an audio interface but as an arrangement: who speaks, in which language, with what authority, and with what opportunity to reply.

2.3 Designing speech systems for under-served languages

A smaller body of work carries these concerns into speech-system development. Reitmaier et al. situate low-resource ASR within the practices of isiXhosa and Marathi speakers, treating community-defined uses, code-switching, data costs, and consent as part of the design problem [25]. They then extend the approach to an unwritten language, building a Gormati spoken-information-retrieval system with a Banjara farming community [23]. Designing for orality, on their account, requires more than adding speech to interfaces organized around written keywords. Hutchinson et al. press further on the categories themselves, questioning whether discrete language identifiers can describe a contact variety at all [11], while Markl and Lai argue that evaluation should be sociolinguistically informed and specific to users and tasks [17].

Ayoka et al. show what happens when a ready-made system enters everyday interaction. Their evaluation of personalized English ASR for people with communication difficulties in Ghana finds that a technically available tool can widen rather than narrow inequities, depending on device access, linguistic fit with Ghanaian English, support, and whether conversation partners adjust to technology-mediated exchange [4]. The setting is communication disability rather than crisis response, but the point generalizes: a speech system reshapes an interaction, and its value cannot be read off accuracy or availability.

2.4 Who governs a deployed system

The same question extends to the organizations that adopt these systems. Bhatnagar et al. argue that responsible humanitarian AI depends on whether organizations retain capacity and authority when technical work is outsourced, and propose co-creation, capacity transfer, and continuing evaluation [8]. Fisher’s humanitarian fieldwork framework makes the companion argument about research relationships, placing reciprocity and value to participants ahead of data collection [9]. For crisis speech technology both bear on who controls voice data and who can change or stop a system once it runs.

2.5 Summary

This literature establishes that trust is relational, that voice is a social arrangement rather than a modality, and that speech systems inherit the conditions of the settings they enter. What it does not yet offer is an account of those

conditions in crisis communication specifically, across more than one language community. That is the gap this paper addresses.

3 Methods

3.1 Study design and sampling

In this study, our aim was to understand how people in ten crisis-affected language communities communicate and receive information during crises, and where voice-based communication might help or exclude them. We pursued this through key informants rather than community members, defining a key informant broadly as someone whose professional or linguistic work gave them direct knowledge of crisis communication in one of these communities. We use *linguistic context* to mean a language community together with its country setting, since one country may host several such communities, and since the community in question is sometimes a regional variety rather than a language as a whole, as with Congolese Swahili. We conducted a qualitative, interpretive secondary analysis of de-identified key informant interview (KII) records from a grant-funded partnership between one U.S. university and an international humanitarian organization, whose broader project seeks to identify language communities for whom speech technologies may be useful in crises. The analysis is secondary in the sense that the first author, who took no part in data collection, analyzed records the second author had gathered for that broader project's own purposes, namely validating quantitative work with informants and producing context briefing notes. The study is exploratory and descriptive, intended to characterize the communication landscape for practitioners and developers working on crisis communication.

Selecting the ten linguistic contexts. We selected the ten study contexts in two steps, with the underlying goal being to identify populous at-risk communities under-served by current language technologies, for which the project remained feasible. We first used a quantitative language-and-crisis-risk index to rank candidate language communities on crisis exposure, language and literacy conditions, and existing digital language support. We then narrowed that ranking by feasibility for the broader project, that is, whether the humanitarian partner held an operational presence and professional network in the location and whether interviews and later voice-data collection could realistically be delivered there. The resulting ten linguistic contexts sit in eight countries across three regions: six in sub-Saharan Africa (Democratic Republic of the Congo, Nigeria, Somalia, and Sudan), three in South and Central Asia (Afghanistan, Bangladesh, and Pakistan), and one in the Caribbean (Haiti). Table 1 pairs each context with its country.

Selecting informants. We sought informants from six groups: academic and linguistic specialists; technology organizations; media and broadcasters; humanitarian and development agencies; community representatives; and government or regulatory authorities. Within each selected context, we recruited through purposive and convenience sampling, drawing on the same partner networks that had made each context feasible to study.

Limits of this sampling. Our study is limited to the contexts where the humanitarian partner already works, and within them to the informants its networks could reach. We therefore do not treat the ten contexts as a representative sample of high-risk language communities. Our study leans toward settings with an established international response presence and away from communities whose crises attract less attention. We acknowledge that these same networks may also have shaped what participants felt free to say about their own organizations. Accordingly, we read each record as one account given from a particular professional position, and we compare accounts across roles and contexts to see where a pattern holds and where it breaks. For the same reason, the number of informants who raise something is not evidence of how common it is, either among practitioners working in these contexts or among the communities they

Table 1. Participants by linguistic context ($n = 35$). Informant categories follow the six the protocol envisioned, of which three were reached: humanitarian covers international and national NGO staff, linguists covers academic and language specialists, and technology covers one technology organization and one social enterprise. Language columns give the language each interview was conducted in. Table 3 matches participant IDs with their corresponding profile: language and informant category.

Language Context	Country	n	Informant category			Interview language	
			Humanitarian	Linguists	Technology	English	French
Congolese Swahili	DR Congo	3	3	0	0	1	2
Dari	Afghanistan	5	3	2	0	5	0
Haitian Creole	Haiti	3	2	1	0	1	2
Kanuri	Nigeria	5	5	0	0	5	0
Lingala	DR Congo	3	3	0	0	0	3
Maay	Somalia	1	1	0	0	1	0
Maxaa	Somalia	3	3	0	0	3	0
Rohingya	Bangladesh	6	4	2	0	6	0
Seraiki	Pakistan	3	3	0	0	3	0
Sudanese Arabic	Sudan	3	1	0	2	3	0
Total		35	28	5	2	28	7

describe. Nevertheless, we believe these limitations do not shadow the importance of this first-of-its-kind large scale study for the context of crisis technologies.

3.2 Participants and data collection

We analyzed records from 35 key informant interviews. Our participants included international NGO staff ($n = 18$), national NGO staff ($n = 10$), linguists ($n = 5$), a technology-sector practitioner ($n = 1$), and a social-enterprise practitioner ($n = 1$). We group these roles into practitioners and language specialists in Table 1, which shows their distribution across contexts. Our sample includes no informants explicitly classified as media, community-representative, or government participants. These roles often depend on knowledge of the target language, so we record each informant’s relationship to that language separately. Most informants spoke the language under study, with thirty documented speakers, two documented non-speakers, and speaker status unrecorded for three. Other demographic and professional attributes, including age, gender, and years of experience, were not consistently documented.

We developed six semi-structured guides tailored to the informant profiles, with contributions from all authors. Our interview protocol covered three areas: the informant’s role and crisis context; how people receive, share, and assess information; and existing voice practices and opportunities for voice technology. We used probes to explore variation by language, literacy, population group, and infrastructure. In several guides, we explicitly asked about trust in information sources and channels. Where relevant, we introduced voice-based tools using AI or automation through examples such as speech recognition, automated call routing, and chatbots. Drawing on the second author’s experience in humanitarian communication, we adapted questions and probes to each informant’s expertise and context.

The second author, who is fluent in English and French, conducted each interview in the informant’s preferred language (Table 1). Interviews took place remotely over Zoom or, where connectivity required it, through WhatsApp voice notes, and lasted 60 to 90 minutes, with participants compensated for their time. The interviewer produced contemporaneous near-verbatim English records, rendering French responses into English while taking notes. We draw quotations from these records; we did not use audio recordings as analytic artifacts.

3.3 Data analysis

The first author led the analysis, following a bottom-up coding process to identify meaningful passages and generate initial codes from participants' accounts. The first author grouped related codes and compared them across linguistic contexts and professional roles to develop preliminary themes concerning communication practices, barriers, and opportunities for voice technology, linking themes to supporting excerpts and documenting divergent accounts that qualified their scope. The analysis underwent two rounds of revision with the third author. We present the themes in the Findings section.

3.4 Researcher positionality

We acknowledge that our experiences and positionality shape how we approach and interpret this research. Our team brought experience in HCI, speech and language technology, humanitarian crisis communication, and work with underserved linguistic communities. The interviews were conducted within the humanitarian partner's professional context, which shaped access and questioning. These perspectives drew our attention to language representation, data control, infrastructure, and organizational practice, while potentially orienting our interpretation toward technical and institutional explanations.

3.5 Ethics Board Approval, Privacy, and Anonymity

This study was approved by the humanitarian partner's ethics review board as well as the university partner's IRB board. Participants signed a consent form with separate choices for audio recording, sharing de-identified records with the university research partner, and quotation in publications, and they could accept or decline each use independently. Before sharing the records, the humanitarian partner removed participant identities and named organizations. We attribute excerpts only by linguistic context, broad professional role, and a sequence number within that context. Because these linguistic and professional communities are small, we review each quoted excerpt and its surrounding detail for re-identification risk before publication.

4 Findings

Across the interviews, crisis messages reached people only when the speaker, language variety, channel, device access, and chance to respond aligned. We call this arrangement a *relational voice infrastructure*. The findings examine three parts of it: how relationships make information credible and answerable; how spoken communication works around language and literacy barriers; and how material access, household control, and automation shape voice technology. Table 2 summarizes this structure.

Table 2. Overview of the themes and subthemes organizing our findings.

Theme	Subthemes
Trusted voices are crisis communication infrastructure	Trust resides in people and relationships; reach is not answerability; sensitive communication needs different routes.
Orality works around structural language and literacy inequalities	Speech is more than a workaround for literacy; institutions produce the oral last mile; writing travels upward while voice travels outward; language fit requires more than a language label.
Voice technology extends participation within material and social limits	Voice technology enters an existing communication ecology; material and household access set the ceiling; automation changes provenance and inclusion.

Supporting excerpts for the participant IDs cited throughout the Findings are grouped by theme and subtheme in Appendix B, in the same order in which they appear within each subtheme.

4.1 Trusted voices are crisis communication infrastructure

4.1.1 *Trust resides in people and relationships.* Credibility depended on a fit between the messenger, the task, the language, and the form in which that messenger normally speaks, and this held across the Dari, Haitian Creole, Maay, Maxaa, and Sudanese Arabic accounts. In the Dari context, mosque speakers were trusted orally, whereas written communication was considered more reliable for institutional information (L02-INGO-01). Authority also depended on the topic. A Haitian Creole participant distinguished the meteorologist whose voice carried authority for weather alerts from the doctor whose voice carried authority for cholera and other health information:

“There’s a well known meteorologist that we don’t normally listen to. But when she talks about rain, then we listen. If I hear an audio message with her voice, it means something to me. But if I hear ... a sports broadcaster telling me it will rain, it means less to me.” (L03-INGO-01)

Trust did not transfer automatically from a person to every topic, or from a trusted institution to every format. Social proximity also mattered. Maay minority-rights volunteers built trust by speaking the language and dialects of the communities they served (L06-INGO-01), while Maxaa participants described imams and elders as familiar links to institutional information (L07-NNGO-01). In Sudan, radio stations sometimes re-recorded humanitarian voice notes in their own voices (L10-NNGO-01). Hearing the message from a familiar station voice made it easier to accept. But the organization that wrote it was then no longer the one speaking, and could not be sure the wording survived.

4.1.2 *Reach is not answerability.* Participants distinguished four functions that a single channel may not provide together: reaching an audience, being accepted, allowing a response, and preserving the message. The Haitian and Maay accounts show that reach itself had geographic and social dimensions. Radio could travel beyond the range of a local megaphone, yet its coverage remained uneven; community leaders could provide a socially accepted route without ensuring that residents found the message convincing (L03-INGO-01; L06-INGO-01). Wider distribution and messenger legitimacy increased the chance that information would be encountered, but neither gave recipients a way to question it or obtain a response.

A few organizations built the return path that was missing. A Lingala hotline connected callers directly to human operators, while a Rohingya audio program carried community concerns to service providers and returned with their

responses (L05-INGO-02; L08-INGO-02). One exchange was immediate and the other took time, but in both cases somebody was responsible for receiving a question and coming back with an answer. The Lingala participant described what that made possible:

“We can ask questions, and the person on the other side can answer. ... When there are crises, there are people who are hired to be hotline operators so you can call a Green Line and they answer your questions instantly.” (L05-INGO-02)

What made these channels answerable was not the audio. It was staffing. Somebody was paid to pick up, find out, and come back with an answer. Passing a message through people also changes it, though. A Kanuri programme addressed this by recording standardized messages so field staff could pass them on without “twisting it through multiple channels” (L04-INGO-01). Building a system people can answer back to therefore means budgeting for the people who will do the answering, and deciding in advance how much a message may change on its way through them.

4.1.3 Sensitive communication needs different routes. Across seven contexts, participants chose a route by asking who else would hear, not by asking which format was easiest. What made a channel private was the social arrangement around it, not any property of the technology.

A Maxaa consultant used voice notes for everyday client work, because “if you don’t know how to read and write, you use voice,” but moved to Zoom “because I don’t want other people to listen” (L07-NNGO-01). A Congolese Swahili participant drew the same line in the opposite direction, reserving speech for public announcements from a recognized official and noting that “for confidential or personal messages, the preference is to do it by writing” (L01-INGO-02). These are opposite choices of medium made on the same grounds, since each participant was asking who else would hear or read the message rather than which format was easier. The written route carried its own restriction, serving only those who “know how to read or write.”

A recording could also change status without being edited. In Seraiki programming, an audio message sent to someone’s personal WhatsApp was also played “collectively via loudspeaker” at meetings (L09-NNGO-01), so only the room and the person controlling playback had changed. The same shift happened at platform scale in Sudan, where reliable information retreated from a disinformation-filled Facebook into WhatsApp groups that were “more organic but restricted access based on trust” (L10-SE-01).

Two accounts show what happens when the route does not fit the message. A Dari participant described the same communities giving opposite answers depending on how they were asked: over a hotline they distrusted the leaders who represent them, while in face-to-face collection conducted with those leaders present they preferred information to come through them:

“In front of their community leaders, out of fear of retaliation or future aid opportunities, they will not disclose this.” (L02-INGO-02)

In Kanuri engagements, a woman and male youth asked directly for “a phone or hotline” offering “an anonymous way of making his or her complaints” (L04-INGO-03). The same community leader who makes a flood warning believable is the person a complainant most needs to avoid, so a service that offers only the trusted community channel will never hear what people are afraid to say in front of their leaders.

Adding a confidential channel runs into a limit these accounts also record. A Rohingya linguist explained that unmarried girls are not permitted phones of their own but “can use their parents’ phone to call people and can talk through their parents phone” (L08-LING-02), so their access is supervised by design and a service built around private

personal use would be unavailable to the people whose accounts are hardest to obtain. Closing a channel protects some people and excludes others, and which effect dominates depends on who already holds a device and a social network.

4.2 Orality works around structural language and literacy inequalities

4.2.1 Speech is more than a workaround for literacy. A Seraiki informant who writes comfortably in Urdu and English, but who never learned to write the language they grew up speaking, described what that language is to them: “I think and breathe in Seraiki. ... My dreams are always in Seraiki” (L09-NNGO-03). Speech was not what remained once reading failed. It was how they used the language they identify with most, while the languages they could write were the ones that belonged to school and work. Haitian Creole participants described speaking and listening in similar terms, as the familiar way of engaging with information rather than a fallback (L03-INGO-02). Literacy was specific to a language and an activity, not a stable attribute that divided participants into literate and non-literate groups.

The same distinction appeared in Lingala, where an informant described people who speak and understand the language fluently yet struggle to read it aloud or write it, including those with formal education:

“People speak it well, but even educated people have difficulty reading and writing Lingala. ... They listen.

They speak. But when it comes to reading and writing, they have difficulties.” (L05-INGO-01)

This relationship was not uniform across settings or situations. In rural Lingala-speaking areas, text could travel more reliably than voice notes over an unstable network (L05-INGO-02). A preference for speech was not a fixed trait of a population. The same speaker might choose a voice note at home, a text message when the network was weak, and a written language they had been schooled in for work.

4.2.2 Institutions produce the oral last mile. Participants connected these differences in oral and written competence to institutional choices. A Lingala informant traced them to an education system inherited from Belgian rule, in which the first two years of primary school are taught in Lingala before, “in the third year, they start teaching in French.” The informant made the comparison explicit: “Lingala is like French in France, English in England, English in the US. Lingala is that to us. But education is provided in French” (L05-INGO-01). A Haitian Creole participant described the same transition from the child’s side, entering a school where “everything is forced in French” and being “punished because I cannot express myself in French,” with parents who “cannot support me” in it (L03-INGO-02). In both accounts people end up fluent in a language nobody taught them to write.

The same hierarchy runs from education into humanitarian communication. Somalia’s constitution names both Maay and Maxaa, yet a Maay participant reported that “all the institutions working in education don’t publish books and writing materials in Maay. Only in Maxaa” (L06-INGO-01). In Kanuri-speaking areas, “most of the radio programming is in Hausa,” which leaves community leaders to “listen to news, often in other languages, and translate and interpret into the Kanuri language” (L04-INGO-01). The interpreting step is unpaid, unrecorded, and performed by whoever happens to have a radio.

Across these contexts, the oral “last mile” was produced by an institutional chain: schools, media, and organizations privileged one language in durable materials, while community members carried the resulting information into another language through speech. People were speaking because a ministry, a newsroom, or a headquarters had already chosen a language the listener does not read, which is a different problem from a listener who cannot read.

4.2.3 Writing travels upward while voice travels outward. The communication mode an organization preferred depended on which way the message was travelling. A Maxaa participant put both directions in a single sentence: “writing is

important for people giving funds to us. The voice we use only with our clients ... And written we use with other stakeholders" (L07-NNGO-01). The same division appeared in Seraiki-speaking areas, this time with the languages attached. Material from government and humanitarian organizations arrived in Urdu, and "sometimes, they deliver information, like brochures, banners, etc., written in English," even though "most people in rural areas are illiterate and don't even speak Urdu or English. They only understand their local language." Local organizations then carried that content outward through mosque loudspeakers and community sessions, where "discussions will always be done at the community level in the local language" (L09-NNGO-01), so that the brochure and the announcement came from different people, in different languages, for different audiences.

Writing and voice were not always split along the same line. In Haiti the division ran between formal and informal rather than between upward and outward. A participant counted "about 150 radios, but only 2 print newspaper" on FM as evidence of an oral culture (L03-INGO-01). Even so, the same participant called a voice note sent to a group "not a formal method," a status reserved for social media, and set radio apart from both as what people turn to "when a catastrophe is hitting." Speech was both the least formal channel available and, in an emergency, the most dependable one. Connectivity shaped the choice as well, and a poor network could remove it altogether. In rural Lingala-speaking areas, "network is unstable and people with poor connection will be able to receive written messages more easily," while voice notes were used in peri-urban areas where the internet was better (L05-INGO-02).

What is uneven across these accounts is therefore not which medium is better, but who does the converting. Staff and volunteers turned written material into speech on the way out to communities and turned what they heard back into writing on the way up to funders, as the Maxaa participant describes doing both, so that at each end the organization decided what form the message took.

4.2.4 Language fit requires more than a language label. Providing a named language did not by itself make communication locally appropriate. A Congolese Swahili participant distinguished the variety they speak from the Swahili of Nairobi, Uganda, and Tanzania, and warned that the difference produces mistranslation (L01-INGO-02). In Kanuri-speaking areas the same problem arose inside a single language, where translations are produced mainly in the Yirwa and Manga dialects while many crisis-affected speakers use a mixed Kanuri shaped by displacement (L04-INGO-01).

A Rohingya linguist tested a machine translation system and traced its errors to the data behind it:

"Google Gemini translates Rohingya language quite impressively. Of course, there are a number of inaccuracies ... because the data they have collected was mostly from NGOs, the problem is that NGOs worked with Chittagonian speakers or non-native Rohingya speakers ... But when I tried to translate my poems into English, it happened remarkably well. But a number of things are wrong, and this comes down to inaccurate data and the data collection." (L08-LING-01)

What worked and what broke, on this account, depended not on the system but on which speakers its training data came from. The Congolese Swahili and Kanuri accounts make the same point about variety, and in all three a system could list the language as supported without having been built from the language people actually use.

A Sudanese Arabic participant named the requirement as localization rather than translation, "making the information more connected, more relevant ... giving a sense of ownership" (L10-NNGO-01). A system can therefore list Rohingya or Swahili as a supported language and still be built from the wrong speakers of it. These accounts set a more specific condition than language coverage: the local variety, recorded by people from the community, and checked by the people who will have to act on what it says.

4.3 Voice technology extends participation within material and social limits

4.3.1 Voice technology enters an existing communication ecology. The channels described so far are the ones any new voice system would join, and participants judged such a system by whether it reached the people those channels currently miss. Some of those channels are already spoken, and communities run them without any organization's involvement. In the Rohingya camps, where fire is a constant threat, people warn each other by recording and forwarding WhatsApp voice clips:

“if there's a fire in Camp [1], instantly if a Mahji or someone needs to pass information, they will send a WhatsApp voice clip to Camp 2 to alert them of possible fire spread. Refugees will send voice clips instantly during emergencies. This is very impactful, but there are no voice methods to reach people without WhatsApp.” (L08-INGO-01)

The limit is not the tool but the people outside it, those without WhatsApp, whom the camp's fastest channel cannot reach. Seraiki organizations get around this by taking the audio off the phone and playing one recording aloud at meetings, where people “listen to the audio collectively via loudspeaker” (L09-NNGO-01).

When we asked what voice technology might offer, participants across the ten contexts described a return path rather than better delivery. A Kanuri participant expected “a very huge impact, especially if it is two-way in form, allowing people to also share information,” tying this to “the literacy levels and how timely the information comes in” (L04-INGO-02). A Dari linguist described speak-in, speak-out interaction as the clearest way around a literacy requirement at both ends of an exchange (L02-LING-01), and a Rohingya participant wanted automatic Rohingya-to-English translation so that young people could “properly express their interests, and build relationships with the international community” (L08-INGO-03). The shift these answers describe is one of role, from an audience that receives warnings to people who can say something back. It asks of a system what Section 4.1 found organizations providing through staffing.

Others named the particular people their channels were missing. A Haitian participant valued radio for its reach but wanted a way to “target specific groups,” because “with the radio, we don't know who it reaches” (L03-INGO-01). Where devices and digital skills were scarce, a Dari participant suggested that organizations “set aside a cellphone that is free” with numbers connecting callers to named agents, while warning that “voice technology would not be useful until other barriers, like digital education and network connectivity, are addressed” (L02-INGO-01). Neither proposal turns on what a system can do with speech. Both are about who gets reached, and both would need an organization to set them up.

4.3.2 Material and household access set the ceiling. Every channel that carries voice depends on electricity, a network, data, and a working device, and each of these can fail at the moment it is needed. A Haitian participant estimated that “about 60% of the population doesn't have access to electricity or isn't connected to the electrical grid,” and observed that “internet functions very poorly during hurricanes” (L03-INGO-01). A Seraiki participant extended that failure well beyond the event itself:

“In the flood areas, they don't have energy for months after floods. So, even during the post-disaster period ... there's a communication blackout.” (L09-NNGO-03)

How many people have a signal before a crisis therefore says little about how many can be reached during one, because what matters is whether the power and the network are still working when the warning goes out.

Where the infrastructure held, the household could still stop a message. A Kanuri participant described the single household phone as the male head's, since "if there is only one phone in the home, it will primarily be the man who uses the phone" (L04-INGO-02), and a Maay participant set out the whole chain:

"those having the most gap in terms of accessing information are women, children, and older people. And people with disabilities also ... 1) men access information because they own the phone ... 2) men lead the family and men decide how information is shared within the family; 3) women's education is very rare, especially in rural communities; 4) older people and people with disabilities don't have telephones." (L06-INGO-01)

In Seraiki riverine communities the same arrangement produced delay rather than refusal: "we see one smartphone per household. And it is in possession of the male family member. So, if he is busy outside, he will only convey the message to other family members when he gets home. Or he might forget to share the message" (L09-NNGO-02). This is the same arrangement that made a confidential route unavailable to the Rohingya girls in Section 4.1, seen from the delivery side rather than the feedback side. Counting phones in a household says little about who can be reached or when, and a voice service that removes the need to read and write leaves the question of who holds the handset untouched.

4.3.3 Automation changes how a voice is judged and who it reaches. Automation changes more than how a message is produced, altering what a listener can tell about who is speaking, whether a community will accept a machine voice at all, and who an audio-only interface excludes. A Haitian participant described an AI-generated voice note that circulated during a cold spell, in which a voice announced that "the end of times has come, it's snowing in Haiti":

"She believed it and didn't realize it was AI ... often we struggle to verify what is shared on social media. But for my parents, everything that is on YouTube is true." (L03-INGO-01)

The clip arrived in the form people already trust and carried no sign of where it came from. If a recognizable voice is how listeners decide whether to believe a warning (Section 4.1), a system that speaks in a synthetic one would need to supply some other way of showing who is speaking and on whose authority.

A Lingala participant showed that acceptance is settled socially rather than individually, reporting that "there are people who think AI is sorcery. How can a machine know everything?" and describing how the view travels between generations, since "a grandpa can tell you that AI is sorcery and as a youth, with access to modernity, you will say 'my grandpa says it's sorcery'" (L05-INGO-02). A younger person might accept the technology and still defer in front of an elder, so willingness expressed by an individual is a poor guide to whether a system will be used.

Inclusion was uneven as well, and a Kanuri participant set out the gradient:

"Those who are blind can still access radio. Kanuri-speakers who are deaf or mute struggle more; sign language is not fully adapted here, so it's more difficult for them to access information. Sign language would be possible, but they would need to be taught sign language first." (L04-INGO-02)

A Rohingya linguist reached the same limit from the opposite direction, judging audio "a perfect solution" for a community that prefers oral communication almost universally, and then returning to people with hearing disabilities, who "don't have access to sign language or things like that" (L08-LING-02). Automation therefore removes the cue by which listeners judged who was speaking, while a voice interface chosen for its accessibility still excludes those who cannot hear it.

5 Discussion

In the preceding sections we have presented how key informants across ten linguistic contexts in eight countries describe crisis communication in the communities they work with. Credibility attached to particular speakers and tasks rather than to channels, speech absorbed inequalities that institutions had created elsewhere, and the usefulness of a voice system was set by conditions around it. Our contribution is to name the five conditions that had to hold for a message to arrive, and to show that they fail separately, so that whether a spoken message reaches someone depends on an arrangement in which the technology is one part. What mattered was never whether a system could deliver a spoken message, but whose voice would carry it, who could answer back, and who would still be missed.

Contexts differ, and often contradict each other. The ten contexts do not converge on a single arrangement, and what serves people in one can be the opposite of what serves them in another. These are not inconsistencies to be resolved by better evidence but the shape of the problem, and a design that suits one context can fail in the next for reasons no research community can address alone.

A problem this complex needs collaboration across fields and institutions. Making a speech system work in these settings would need crisis responders, government authorities, broadcasters, and the affected communities themselves, alongside the researchers who build it. That is not a reason to withdraw, but a reason to be specific about which part of the problem a system can address and to state its limitations openly.

Speech technology and HCI have largely approached this separately. Research on low-resource speech technology has concentrated on models, datasets, and benchmarks for languages such as those we studied [1, 12, 17], while HCI research on crisis and displacement has concentrated on communication practice, trust, and access with little reference to speech systems [5, 10, 29]. Few studies treat crisis communication in these languages as the condition a speech system would have to meet rather than as a domain to apply one to, though a small body of work joins the two [4, 14, 23]. Our account is offered to both communities, and with it a case for supporting the practitioners and linguists who already hold this knowledge to decide what is worth building. The subsections below situate the findings against that literature and suggest possible directions for future work.

5.1 Relationships and Staffing Turn Reach into Answerability

This community has established that crisis information is produced through social and institutional arrangements rather than delivered through neutral channels. Across crisis informatics and humanitarian HCI the recurring finding is that whether information reaches someone depends on relationships, institutions, and infrastructure rather than on a channel being available [21, 28, 29], and work in displacement and humanitarian settings reaches the same conclusion about trust and participation [2, 13, 30]. Much of that evidence comes from platform traces, while the channels our participants treated as decisive leave almost none: a chief with a microphone, a radio bulletin interpreted aloud into another language, a voice note replayed to a room. Interviews reach what no trace records.

What our work adds is which part of an arrangement fails, and that this can be compared across settings. Five conditions recurred: a speaker whose authority fits the task, a language and variety the listener uses, a channel that survives the crisis, a device the listener controls, and an organization that has assigned someone to answer. They fail separately, as when the Maay feedback loop failed on language while succeeding on staffing. The last is the condition technology most easily hides, since digital feedback can give organizations the appearance of accountability while nothing changes for the people who spoke [16]. In the cases we saw, people could get an answer only where an

organization had paid someone to give one. Answerability is therefore a budget line before it is a design feature, and the question to ask of a system is not whether it collects feedback but who is on duty to reply.

5.2 Language Coverage Does Not Ensure Fit or Trust

Speech research now covers these languages at a scale that was out of reach a decade ago. Massively multilingual models, crowdsourced corpora, and self-supervised encoders list support for thousands of languages, measured as error rate on shared benchmarks [1, 3, 19], and the field is already turning to how speech data is collected [27].

That progress is what makes a further distinction worth drawing. A thousand supported languages, or a falling error rate, is easily read as a claim about who can now be reached, while our participants describe conditions under which such a system would still fail for many people inside that count. It is more useful to say which variety a system supports, for which speakers, and for what task, and to evaluate it with the people expected to act on what it says [17, 24]. Language identifiers are too coarse to carry that claim [11], and our Congolese Swahili and Rohingya accounts extend it to provenance, since a system may list a language while having learned it from the wrong speakers of it.

Human-centred work shows who pays when a count overstates reach. Speakers outside the assumed variety carry the cost themselves, hyper-articulating or modifying an accent, and often reading the breakdown as a personal failing rather than a design one [15, 32], while different kinds of failure damage trust differently [7]. A Congolese Swahili participant described systems demanding repeated input, which in a crisis costs time the speaker may not have.

Synthetic speech inherits this and adds one problem of its own. A generated voice carries a variety whether or not anyone chose it [18], and listeners identify synthetic audio poorly [6]. Participants in our study judged a warning by recognizing the voice that carried it. A synthetic voice removes that cue without announcing that it has, so a system using one has to show where the message came from by some other means.

5.3 The Five Conditions Require Shared Responsibility

Beyond the ten settings studied here, what this work offers is a way of stating what a crisis voice system has to satisfy: who speaks, in which variety, over what channel, on whose device, and with whom answering. These conditions are separable, and each can be checked before a system is deployed rather than inferred after it fails.

Their difficulty is that they do not divide along the lines the two fields work on. Design research has established the value of a private channel for sensitive questions [22] without having to ask what happens when the user does not control the phone, and removing written assumptions from a speech interface [23, 25] leaves untouched the schools, broadcasters, and donors that put them there. Neither is an oversight, since questions of this kind sit outside the usual scope of both fields, which is also why they have gone unasked.

Deciding which of these to test is not a question for researchers alone. The humanitarian staff, linguists, and community organizations in this study hold the knowledge that would settle most of them, and the most useful contribution speech and HCI research can make is to work alongside them: agreeing together what is worth building, what would count as it working, and who remains responsible once it is running.

6 Limitations

Our participants were humanitarian practitioners and language specialists, not crisis-affected community members recruited as such. Their accounts carry professional observation and institutional knowledge, and much of what we report, including what people feared to say in front of their leaders, how girls reach a shared phone, and how a warning

is judged, is described on others' behalf. These are informed accounts, but they are not first-hand ones, and the claims in this paper need checking with crisis-affected speakers directly.

The ten contexts and the informants within them were reached through one humanitarian organization's networks, which bounds the range of institutions and critical perspectives represented and tilts the sample toward settings with an established international response presence (Section 3.1). We did not conduct co-design or member checking with affected communities, and we neither built nor deployed a system, so nothing here shows how one would perform in these settings.

The form of the records also bounds what can be claimed. They are contemporaneous notes taken by the second author, who conducted every interview and is fluent in both languages, and we treat them as a reliable account of what was said, including for the seven interviews held in French and written down in English as they went. Because no interview was transcribed from audio, however, the records support thematic interpretation but not close claims about word choice, hesitation, or interaction. Attributes beyond professional role and relationship to the target language were not documented consistently enough to report.

Finally, the analysis was led by one researcher and revised with a second rather than coded independently by two, so we report no measure of inter-coder agreement. We treat coding here as interpretive work rather than measurement, and the two rounds of revision surfaced disagreements that reshaped several themes, but a reader should weigh the themes as one considered reading of the records rather than as a result reproduced independently.

7 Conclusion

We analysed 35 key informant interviews on crisis communication across ten linguistic contexts in eight countries, and found that a spoken message reached someone only when five conditions held together: a speaker whose authority fitted the task, a language and variety the listener used, a channel that survived the crisis, a device the listener controlled, and an organization that had assigned someone to answer. These conditions fail separately, and meeting four of them is not a partial success, since a message that satisfies every condition but one does not arrive.

Stating them separately makes them checkable, and it shows that most of them lie beyond what better speech technology can settle. They belong as much to crisis responders, broadcasters, and government authorities as to the researchers who build these systems, and the practitioners and linguists in this study already hold much of what would answer them.

What would settle these questions is work carried out with the people whose communication they describe, in studies that speech researchers, HCI researchers, and those practitioners run together rather than in turn.

References

- [1] Jesujoba O. Alabi, Xuechen Liu, Dietrich Klakow, and Junichi Yamagishi. 2025. AfriHuBERT: A self-supervised speech representation model for African languages. In *Interspeech 2025*, 4023–4027. doi:10.21437/interspeech.2025-1437
- [2] Asam Hamed Abbas Almohamed, Reem Talhouk, and Dhaval Vyas. 2022. Towards a Conceptual Framework for Understanding the Challenges in Refugee Re-settlement. *Proc. ACM Hum. Comput. Interact.* 6, GROUP, Article 37 (2022), 27 pages. doi:10.1145/3492856
- [3] Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2020. Common Voice: A Massively-Multilingual Speech Corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 4218–4222. <https://aclanthology.org/2020.lrec-1.520/>
- [4] Gifty Ayoka, Giulia Barbareschi, Richard Cave, and Catherine Holloway. 2024. Enhancing Communication Equity: Evaluation of an Automated Speech Recognition Application in Ghana. In *International Conference on Human Factors in Computing Systems*. *International Conference on Human Factors in Computing Systems*, 1–16. doi:10.1145/3613904.3641903
- [5] Abdul Aziz and Karen E. Fisher. 2026. Mediated information grounds: living with(in) pareshani, social networks and gendered media practices among the Rohingya refugees. *Information, Communication & Society* (2026), 1–20. doi:10.1080/1369118X.2026.2685119

- [6] Sarah Barrington, Emily A. Cooper, and Hany Farid. 2024. People are poorly equipped to detect AI-powered voice clones. *Scientific Reports* 15, 1 (2024), 11004. doi:10.1038/s41598-025-94170-3
- [7] Amanda Baughan, Xuezhong Wang, Ariel Liu, Allison Mercurio, Jilin Chen, and Xiao Ma. 2023. A Mixed-Methods Approach to Understanding User Trust after Voice Assistant Failures, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–16. doi:10.1145/3544548.3581152
- [8] Tigmanshu Bhatnagar, Maaryia Omar, Davor Orlic, James Smith, Catherine Holloway, and Maria Kett. 2025. Bridging AI and Humanitarianism: An HCI-Informed Framework for Responsible AI Adoption, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–17. doi:10.1145/3706598.3713184
- [9] Karen E. Fisher. 2022. People First, Data Second: A Humanitarian Research Framework for Fieldwork with Refugees by War Zones. *Comput. Support. Cooperative Work*. 31, 2 (2022), 237–297. doi:10.1007/s10606-022-09425-8
- [10] F. Hussain, S. N. Amin, and T. Trussler. 2026. Rohingya participatory media in protracted displacement: Counterpublic formation through calibrated participation. *Media, Culture & Society* (2026). doi:10.1177/01634437261458549
- [11] Ben Hutchinson, Celeste Rodríguez Louro, G. Collard, and Ned Cooper. 2025. Designing Speech Technologies for Australian Aboriginal English: Opportunities, Risks and Participation, In Conference on Fairness, Accountability and Transparency. *Conference on Fairness, Accountability and Transparency*, 108–124. doi:10.1145/3715275.3732010
- [12] Sukairaj Hafiz Imam, Babangida Sani, Dawit Ketema Gete, Bedru Yimam Ahamed, Ibrahim Said Ahmad, Idris Abdulmumin, Seid Muhie Yimam, Muhammad Yahuza Bello, and Shamsuddeen Hassan Muhammad. 2025. Automatic Speech Recognition for African Low-Resource Languages: Challenges and Future Directions, In Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025). *Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025)*, 89–94. doi:10.18653/v1/2025.africanlp-1.13
- [13] Carlos Augusto Bautista Isaza, D. Tatar, Aakash Gautam, Pedro Reynolds-Cuellar, and Sang Won Lee. 2026. The Precarity of Trust: Understanding Tensions in Information Systems for Humanitarian De-Mining in Rural Colombia. *ACM Transactions on Computer-Human Interaction* (2026). doi:10.1145/3829350
- [14] Md. Adnanul Islam, Patrick Olivier, and Delvin Varghese. 2024. Leaving No One Behind: An Agenda for Voice-based Engagement of Remote Communities, In International Conference on Information and Communication Technologies and Development. *International Conference on Information and Communication Technologies and Development*, 175–191. doi:10.1145/3700794.3700812
- [15] Siyu Liang and Alicia Beckford Wassink. 2026. “This Wasn’t Made for Me”: Recentering User Experience and Emotional Impact in the Evaluation of ASR Bias. 7308-7326 pages. arXiv:2604.21148. doi:10.1145/3805689.3806448
- [16] Mirca Madianou, J. Ong, Liezel C. Longboan, and Jayeel S. Cornelio. 2016. The Appearance of Accountability: Communication Technologies and Power Asymmetries in Humanitarian Aid and Disaster Recovery. *Journal of Communication* 66, 6 (2016), 960–981. doi:10.1111/jcom.12258
- [17] Nina Markl and Catherine Lai. 2021. Context-sensitive evaluation of automatic speech recognition: considering user experience and language variation, In HCINLP. *HCINLP*.
- [18] Shira Michel, Sufi Kaur, Sarah Elizabeth Gillespie, Jeffrey Gleason, Christo Wilson, and Avijit Ghosh. 2025. “It’s not a representation of me”: Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services, In Conference on Fairness, Accountability and Transparency. *Conference on Fairness, Accountability and Transparency*, 228–245. doi:10.1145/3715275.3732018
- [19] Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2024. Scaling Speech Technology to 1,000+ Languages. *Journal of Machine Learning Research* 25, 97 (2024), 1–52. <https://www.jmlr.org/papers/v25/23-1318.html>
- [20] Hemant Purohit, Cody Buntain, Amanda Lee Hughes, Steve Peterson, Valerio Lorini, and Carlos Castillo. 2025. Engage and Mobilize! Understanding Evolving Patterns of Social Media Usage in Emergency Management. *Proc. ACM Hum. Comput. Interact.* 9, 2, Article CSCW067 (2025), 39 pages. doi:10.1145/3710965
- [21] Marianne Aubin Le Quere, T. Chiang, K. Levy, and Mor Naaman. 2022. Information Needs of Essential Workers During the COVID-19 Pandemic. *Proc. ACM Hum. Comput. Interact.* 6, CSCW2, Article 306 (2022), 19 pages. doi:10.1145/3555197
- [22] Pragnya Ramjee, Mehak Chhokar, Bhuvan Sachdeva, Mahendra Meena, Hamid Abdullah, Aditya Vashistha, Ruchit Nagar, and Mohit Jain. 2024. ASHABot: An LLM-Powered Chatbot to Support the Informational Needs of Community Health Workers, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–22. doi:10.1145/3706598.3713680
- [23] Thomas Reitmaier, Dani Kalarikalayil Raju, Ondrej Klejch, Electra Wallington, Nina Markl, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. 2024. Cultivating Spoken Language Technologies for Unwritten Languages, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–17. doi:10.1145/3613904.3642026
- [24] Thomas Reitmaier, Electra Wallington, Ondrej Klejch, Nina Markl, Léa-Marie Lam-Yee-Mui, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. 2023. Situating Automatic Speech Recognition Development within Communities of Under-heard Language Speakers, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–17. doi:10.1145/3544548.3581385
- [25] Thomas Reitmaier, E. Wallington, D. Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, P. Bell, and Simon Robinson. 2022. Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers, In International Conference on Human Factors in Computing Systems. *International Conference on Human Factors in Computing Systems*, 1–17. doi:10.1145/3491102.3517639