

Situating Asynchronous Voice in Rural Africa

Nicola J. Bidwell^{1,2} and Masbulele Jay Siya³

¹Department of Informatics, University of Pretoria

²CSIR: Meraka Institute

³Mankosi AA, Eastern Cape, 5140, South Africa

nic.bidwell@gmail.com, jaysiya26@gmail.com

Abstract. Designing for oral users in economically poor places has intensified efforts to develop platforms for asynchronous voice. Often these aim to assist users in rural areas where literacy is lowest, but there are few empirical studies and design tends to be oriented by theory that contrasts the mental functions of oral and literate users, rather than by local practices in social situations. We describe designing an Audio Repository (AR) based on practices, priorities and phone-use in rural Africa. The AR enables users to record, store and share voice files on a shared tablet and via their own cell-phones. We deployed the AR for 10 months in rural Africa and illiterate elders, who have few ways to use free or low-cost phone services, used it to record meetings. Use of, and interactions with, the AR informed the design of a new prototype. They also sensitized us to qualities of collective sense-making that can inspire new interactions but that guidelines for oral users overlook; such as the fusion of meaning and sound and the tuning of speech and bodily movement. Thus, we claim that situating design in local ways of saying enriches the potential for asynchronous voice.

Keywords: Oral users, Rural Africa, Asynchronous voice, Social media.

1 Introduction

Most of the world's 800 million illiterate adults live in economically poor regions; in fact, interpolating Worldbank indicators, in Africa fewer people can read and write than there are cell-phone subscriptions [23]. Cell-phones offer many new opportunities to people in technology-sparse regions but they may also amplify gaps between oral and literate users and this has motivated recent efforts to develop *asynchronous voice* platforms for rural settings where literacy is lowest. Such efforts include interactive voice applications and services for cell-phones to record and access content produced locally, or by authorities or NGOs [19, 27, 31, 39] or participate in citizen journalism or online social networks [25, 29]. However, there are few empirical studies [29] and design tends to be oriented by theory that contrasts the mental functions of oral and literate people, not by practices in social situations. In this paper we describe a system we deployed in rural South Africa to enable users to record, store and share voice files for free. We aim to show that situating design in local ways of saying can inspire new interactions for asynchronous voice; thus, we start by claiming that theory currently orienting design for oral users omits factors that can prompt ideation.

1.1 Design Principles for Oral Users

Most design guidelines for oral users in economically poor regions [e.g. 13, 32, 37] draw on Ong's theory about cognition and the internalisation of writing [28]. Ong (1982) claimed that literate people perceive words as reified things with an existence and meaning that is separate from speech sounds and this affects memory, categorization and verbal-explanation [28]. Thus, guidelines often contrast the formulaic, prosaic and experiential thinking of oral users with the objective, logical and analytic skills of literate users [e.g. 37]. However, many studies show that literacy alone is not responsible for great shifts in mental functioning. Extensive experimental studies and ethnographic and survey-based descriptions of literacy find that certain writing systems and certain reading and writing activities foster certain forms of thinking [e.g. 36]. For instance, Qur'anic literacy improves performance on some memory tasks and literacy in a certain Liberian script enhances phonological discrimination [36]. These, and other studies [15], also show the equally specific effects of schooling on verbal explanation. Further, while guidelines emphasise oral users' deficits in abstract thinking [37], studies in Africa show that oratures involve abstract and other complex structures [1, 17], harmonic song and polyrhythm execute intricate patterns and that polylingualism, common in Africa, fosters a sensitivity to how a language works [10].

Ong's theory appeals to perspectives that deem thinking to be cognitive, tie learning to information transfer and/or model design problems as tasks. These approaches best suit studying individual capabilities and single-user interactions in controlled settings [14], such as to test oral users' learning from instructional video [20]. They cannot predict interactions between technology, orality, learning and the diverse resources and skills that people use to express and communicate ideas and work with information. Many authors admit that task-focused and cognitive studies do not address the complexity of mundane communication, social factors that effect learning or the politics of evaluating certain skills in certain ways [20]. Yet the epistemic status of these methods promotes design decisions that *are* insulated from real-world messiness. Much design for oral users draws on insights that are framed by specific goals but not by everyday practice, even when informants tell researchers that their main information source is "word of mouth" [e.g. 22]. Indeed, authors can advise grounding design in "practices of community knowledge building and transmission" [37] and "how" as well as "what" people say [32] but then extract requirements, using general-ized models of information transfer, rather than produce ideas within local dialogue.

1.2 Body and Voice

Ingold (2007) argues that Ong's theory inadequately accounts for the ways the printing press and certain literacy pedagogy disconnect graphic inscription from the gesture of writing [16]. For instance, children in the West first encounter letters as tangible objects (e.g. play-bricks) and are drilled to copy their form rather than the gesture that created it [16]. Certainly, cognitive approaches to oral users tend to ignore the actions that produce representational forms. For instance, Medhi, *et al* (2007) suggest

that illiterate people recognize meanings in drawings better than photo-realistic images because they contain less information [21] but not because drawings are traces of hand movements. Interpretations that apply Ong's theory usually conflate print and hand-writing; for example to attribute disruptions in experiential information to writing itself not the printed forms that people write into [e.g. 32]. Overall such analyses separate thought/mind from action/body and neglect the knowing and reasoning constituted by physical manipulation, movement or sensory/motor memory [18].

Relegating bodily abilities to ergonomics deprives design of many constituents of orality. For instance, in Africa, gestures or mnemotechnical devices that attach verbal utterances to places or objects [1, 10] can inspire new ways to support oral practices such as tangible designs that link stories and beadwork [38]. Responding to extra-linguistic and meta-communicative phenomena, such as the mutual tuning of people's voices and bodies even when they talk using phones and cannot see one another [26], may enhance interactions. For instance, it might improve the efficacy of Interactive Voice Response (IVR), which is currently inferior to interacting with a live operator [31]. Recognising embodied social risks and emotions [18] can also guide decisions on interaction modes; for instance, while at ease using IVR to access football updates, users preferred to input to phone keypads in accessing information about HIV [27].

1.3 Designing for Oral Practice

Designing interfaces for asynchronous voice based on principles that *compensate* for what users did not attain via the pedagogies of certain literacies neglects how sense-making is always situated in social contexts. Oral societies use many skills and tools to produce meanings. Mutual fluency in specific oral forms enables members to convey and stabilize information [1] and sustain social bonds that are integral to cohesion and resilience in poor, rural communities. However, design efforts for oral users mainly separate single-user and co-present interactions with technologies such as cell-phones, despite the prevalence of proximate use and intermediary access by users in economically poor areas [e.g. 42]. Work in asynchronous voice is often inspired by the way local radio has contributed to poverty relief by enabling people to share locally relevant information regularly. Most efforts in rural areas aim to support specific interests or domains for social upliftment, such as agriculture, education or healthcare. However, understanding the local communication ecology in which information is shared is vital to avoid amplifying existing gaps in access to technology between genders, generations and areas [e.g. 2, 9, 33]. For instance, women were less able than men to participate in an Interactive Voice Forum (IVF) in Tanzanian farming communities [39]; relatively few farmers given access to an IVF in India conveyed what they learnt to those who were excluded [29]; and, social relations in and around rural clinics contributed to the failure of a healthcare communicator in South Africa [41].

Our design of the Audio Repository (AR), described here, sought to respond to practices and priorities, economics and phone-use in South Africa's rural Eastern Cape. Thus, next, we introduce the setting and our approach to designing to assist local communication. We outline factors that inspired the AR and aspects of its design. Then we summarise use of, and interactions with, the AR and pose some specific and more general implications for designing interactions with asynchronous voice.

2 Setting

Our research was set in Mankosi, along Eastern Cape's Wild Coast. It was prompted by our (Nic's) ethnography in a neighbouring area [3], which shaped the design of a mobile digital storytelling application (Fig. 1) [5] and also exposed difficulties local Xhosa people have in communicating between villages. There are 580 households in Mankosi and inhabitants have extensive kin in twelve villages spread over 30km² of very hilly land. Between 500 and 900 people live in each village, as many temporarily migrate for work or school. Households survive on around \$150 per month, mostly in payments from remote family members and government grants. Up to six adults and eight children live in each homestead, comprised of clusters of thatched, mud-brick rondavals, an occasional tin-roofed 2-room dwelling, animal corrals and a garden for subsistence crops. Hedges or fences around homesteads do not enclose domestic life. Children often move between homesteads and across the day people walk along many paths to communal pasture, forest, taps and dams where they collect water, fire-wood and grasses, graze animals, tend plots and make mud-bricks. There is only one bus a day which takes up to 2-hrs to traverse Mankosi and does not go to all villages.

Like 36% of South Africa's population, inhabitants are governed by a Tribal Authority, which in Mankosi consists of the Headman, a Subheadman in each village and three messengers. The Headman's and Subheadmen's homesteads are sites for local administration, from hosting meetings to notarising proof of residence papers. Post-apartheid Tribal Authorities are separate from other political bodies. Headmen relate by clan to Chiefs and inherit their role patrilineally, but they can replace Subheadmen and permit women to assume these roles. The Headman is paid a government stipend but, like almost all inhabitants, he cannot afford a car or to connect to mains electricity; still, he is, amongst the 20% who can afford to send children to school in a city.

3 Ethnographic Action Research

We learnt about Mankosi's communication ecology and engaged inhabitants in articulating requirements for information sharing and deploying prototypes in Ethnographic Action Research [40]. Critical to our methods, and accountability, are that Jay and eight other Local Researchers (LRs) generated data, advised in design, trialed systems and translated linguistically and culturally; and, we all lived in Mankosi - LRs since birth and Nic for 20 months from 2010 to 2013. We produced data using mixed methods [e.g. 4, 7] including observations, contextual inquiry, interviews, diary studies, auto-ethnography and deploying, and training people to use, different prototypes. Few local people speak English and, while Nic's rudimentary isiXhosa has improved, in most activities Nic spoke in English, others spoke in isiXhosa and LRs translated. We recorded activities in handwritten notes and, sometimes, photos and video.

LRs insisted we consult the Tribal Authority so we began by gaining approval to experiment with communication systems that might assist inhabitants, and perhaps in turn others like them elsewhere. Subsequently, we observed or participated in many community meetings and met with the Tribal Authority almost 40 times to discuss

ideas, plans and problems in designing, trialing and maintaining systems (Fig. 1). Shortly after our first meeting LRs and other people began video-recording conversations, storytelling and presentations. They accumulated over 50 items featuring some 60 people, aged 14 to 80 years, which yielded data on values, priorities and oral style.

We also undertook over 300 interviews, individually or in groups, specifically about communication practices and use of phones, phone services and prototypes. We first interviewed 141 people about phone ownership and use and, as this showed the prevalence of using free Callback services, we interviewed another 16 people about Callback [4]. Then we generated data on the ways people manage mixtures of communication in their daily routines in diary studies across periods of 4 to 10 days. This involved individual or group interviews at the start and end, and 72 short individual interviews that reviewed phone-use in between. We included six male and six female older owners of low-end phones, half of whom are illiterate; and, ten younger, literate owners, some of whom we interviewed using a text-based, chat service that is popular elsewhere in South Africa. A UK researcher, visiting for two weeks, also conducted 17 group or individual interviews with 28 younger feature-phone owners [11].

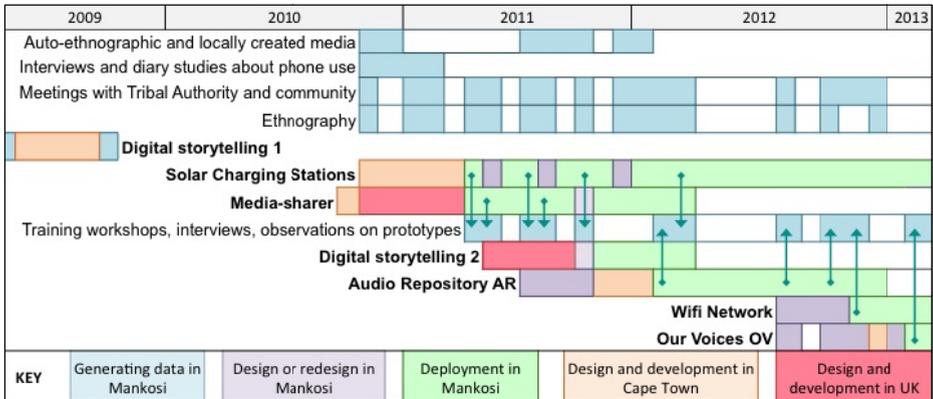


Fig. 1. Timeline of activities to gather data, design, develop and deploy and evaluate prototypes

We shared our data with UK researchers who developed two prototypes, for creating, storing and sharing media. We deployed their first prototype, a media-sharer, in April 2011, which yielded insights for our design of the Audio Repository that we introduced in early 2012 (Fig. 1). The media-sharer ran on two tablets each deployed with solar-powered, cell-phone Charging Stations [7] as we hoped that people would use the media-sharer when they came to charge their own phone for free. We will soon report the design of the media-sharer, which uses real-time, text-based chat between users’ own phones but sends/receives photos later between users’ phones and communally accessible tablets. Inhabitants chose two sites for the Stations that were 2.5Km, or 25-mins walk, apart: in the homesteads of the Headman and a Subheadmen in Mankosi’s poorest area, Ridge [7]. We trained LRs to use the tablet, media-sharer and Stations and then ran seven workshops on the media-sharer with 24 inhabitants, mostly members of the Tribal Authority or Community Association. The media-sharer was

unsuited to most local phones, so we also gave low-end Bluetooth-enabled phones to Tribal Authority and Community Association members. We observed each Station site for over 80-hrs on different days of the week, at different times of day and different times in the year and interviewed 40 people who left or collected phones at Charging Stations [7]. We logged use of the media-sharer automatically; who charged phones each day in notebooks; and, the distance between Stations and the homesteads of 40 Station users. In the ensuing year few of the 700 people who charged phones [7] used the media-sharer. We tried to promote use of the tablet by uploading locally created audio commentaries, photos and videos of local soccer matches and exploring with players how tablets might support the league. In October 2011, UK researchers introduced a digital storytelling prototype for Android smartphones that interfaces with the media-sharer (Fig. 1). They left two phones and a tablet with LRs to loan to people [11], but nobody has used the application yet.

LRs verbally translated audiovisual records and we transcribed audio and/or annotated video. We thematically coded data after each activity and cross-linked themes between activities. We used descriptive statistics to analyse responses to closed questions in interviews and gave written descriptions of varying thickness to UK researchers. We validated and revised interpretations of themes hermeneutically, as dialogue and events produced new meanings, and recently began detailed analyses of video to explore themes that often occurred when we cumulatively integrated codes.

4 Inspiration for Designing the Audio Repository

Our data on co-present communication and phone use in rural Africa confirms, disputes and significantly adds to that already reported in HCI4D [e.g. 4, 6, 7]. Here we present issues that often arose in thematic analyses and oriented designing the Audio Repository. We illustrate some of these issues with ethnographic details.

4.1 Orality and Deliberate Qualities in Speaking

Inhabitants spend much time talking to greet, support, learn, gossip or debate and some 40% of adults, including elders in the Tribal Authority, do not read and write. There are few reading and writing materials locally, even in schools, and people's most common experience of print is a phone's interface. Many recognize the role of orality in local identity and belonging, for instance one young man linked his talent for poetry, inherited from an uncle with whom he performed at ceremonies, to "fighting the westlife" on behalf of the Xhosa. The vast range of video inhabitants recorded focused on the speaker not the visual setting. People said audio recording was a useful feature of phones [4] and some used feature-phones to record local performances, events, conversations, jokes and stories as well as music from the radio [11]. In fact participants in workshops on the media-sharer often recorded themselves speaking or singing rather than take photos. The preferences of men to record their speech and women to record song align with practices in ceremonies and church.

Inhabitants prefer to use phones for voice calls over other services, especially if they are older. For instance, SMS comprised only 1.4% of older people's phone-use

in diary studies. People rarely use free services that send an SMS to recipients who do not pick-up, nor leave voicemail as it consumes airtime they would prefer to use to talk. Despite favoring voice, inhabitants make brief calls as average weekly spending on airtime limits speaking to just 4 minutes at basic rates and few phones can access cheaper data rates for VOIP [4]. This constrains expression, especially as isiXhosa can take three times longer than its English translation, and means phone use is more often planned than impromptu. Inhabitants mostly interact with phones in their homesteads and rarely while walking along paths/roads, in transport or in meetings. They said that prior to owning a cell-phone others phoned or sent a message to someone else in their village who purposively walked to bring them a message or the phone.

Deliberate qualities in speaking also arise in local governance, which is based on customary structures and community meetings. Up to 70 inhabitants (55% male, 45% female, and 60% older than 30 years) attend weekly meetings, in the Headman's homestead, that address Mankosi's internal issues and/or host counselors or municipal officers. Meetings often take 3 hours and can entail walking for over 2-hrs to attend. This limits turnout. People sit according to gender and age to speak a person raises a hand, stands and removes his hat, if he is a man. People attend to the speaker and nobody interrupts. The Tribal Authority aims to reach consensus in meetings and, after opening a meeting, the Headman usually listens silently until the end when he states how an issue would be resolved or discussed further. The Tribal Authority's voluntary secretary writes minutes and attendee's details on paper but does not disseminate these due to cost, infrastructure and because Subheadmen and Headman's messengers are conduits for information across villages. This fuels allegations about the Tribal Authority's inertia, opacity and "forgetfulness". The Tribal Authority said that audio recordings of meetings would "stop people lying", which approximates to being held accountable, and could inform those unable to attend and other inhabitants also hoped communication systems might improve information dissemination.

4.2 Differences in Access to Phones and Literacies in Phone Use

Some 30% of inhabitants over 14 years old do not own a phone, only 25% own feature-phones and people under the age of 35 are twice as likely to own a phone than older people [4]. Older and/or illiterate people often said that they were "not educated to use the phone" or did not use free or low cost services and call rates as they could not "read properly". Indeed, the amount of print in the media-sharer contributed to its lack of use. Younger and/or literate phone owners have many tactics to use text-based services to coordinate communication, seek/offer support or sustain a presence in other's lives. They choose to use these based on their recipients' technological or print literacy [4]. Younger people often use free services and low-cost IRC chat services to communicate with friends and people in cities, including those they do not know. Older people perceive this as playful and irrelevant to local communication. Generational differences in phone-use are acerbated by authority protocols and a dramatic population decline between the ages of 20 and 44 years ('skip generation'). People of parental age are those most likely to have died from illness (often related to

a local HIV prevalence of 29%) or to work in cities and mines. Grandparents often head households but respect for elders can make younger people feel uneasy about teaching them to use cheap services. Indeed, LRs often interacted with phones for elders in workshops. Illiteracy is not stigmatized but Tribal Authority members were shy to reveal that they could not read. The Headman refused to try the media-sharer, though with practice other elders used the tablets' direct manipulation interface easily.

Print and technological literacies have different implications for privacy in phone use. Phone owners said phones helped to make their lives more private and their patterns of calling or walking between homesteads had changed, so now they chose to "shout-out" or make a deliberate missed call ("buzz") depending on their reason to communicate. Those with low literacy gain help from trusted intermediaries and close family members know each other's PIN codes and have various tactics to prevent another's use of their airtime. However, an illiterate mother was concerned that people took photos of others without consent and, meanwhile, we had to discourage LRs, aiming to prevent theft, from saying that shared equipment can track users.

4.3 Temporality and Phone Use

Perceptions of opacity and delay in local governance communication are situated in access to phones and phone services, the pace of life and frequent co-present contact within villages. People spend much time outdoors in Mankosi and an ambient orality pervades daily life as they call between homesteads, sing in gardens, whistle to animals and greet and chat as they walk to communal resources [7]. Most phone owners acquired phones between 6 months and 4 years ago and prioritise using phones to sustain supportive connections [4]. Their phone use tends to split between contact with people living nearby and much farther away. For instance, older participants in diary studies had seen, or predicted to see, nearly a third of the people with whom they had interacted by phone within a day of that interaction; and, another half of their phone interactions were with close family members or friends in places that entail at least a day of travel. Only 13% of older people's phone use was contact with people in other villages in Mankosi and only 10% of their use specifically arranged meeting.

Cost and access to infrastructure affects the temporality of phone use. For instance, some 50% of phone-owners' diaries recorded at least one day when they did not use their phone at all. Firstly, cost limits immediacy in communicating. Most phone-users spend R7 (1.1 US\$) per week on airtime as pay-as-you-go subscribers to MTN's network [4], which entails keeping a small airtime credit. Also inhabitants usually pay on average \$0.70 to charge their phones, often at spazas (small shops) and shebeens (informal bars) [4] and can go for a week without charging so conserve charge by switching off phones [7]. Secondly, cellular network coverage and electricity access limit connectivity. People undertake daily tasks in places with unstable connections and can leave phones to charge for two days if charging is free [7]. Thirdly, low-end phones do not always display that they received calls when they were off. Finally, access to phones or assistance via an intermediary effects temporality; for instance, an illiterate mother always waits for her daughter to walk home from school to read aloud and reply to her messages.

5 The Audio Repository

We designed the AR to run on the 7inch Samsung Galaxy tablets that inhabitants had kept at the Stations, that had charged up to 20 phones a day in the past year. The AR, developed in Java and SQLite, consists of a database and API to make, access and share recordings and create and join groups. We sought to respond to local practices and rapidly deploy a working prototype to allay frustration with the media-sharer. Jay visited the developer, Thomas, in Cape Town. Their design discussions were shaped by Jay’s experiences in trialling the media-sharer, Thomas’ design of our original digital storytelling application [5] and the contrast of the city and Mankosi [34]. Here we outline only the main conceptual, interface and interaction aspects of our design.



Fig. 2. Interfaces to find account (a), record (b), share (c), play or send (d) recordings

All users can record and listen to voice files using the AR on the tablet, whether or not they have an account. However, unregistered users can only interact with files that are made ‘public’. Sharing files does not require owning a phone, unlike for the media-sharer, but users can send/receive files to/from cell-phones over Bluetooth (Fig. 2d). We tested the AR with Nokia 2730 and 2330 Classic phones, since some 60% of local phones are Nokia [4], but it also works for other low-end Bluetooth phones with enough unused storage space. Recording functions are merged in one button, beneath a cassette tape icon, so when users press the red circle, to record, it turns into a black square to stop. The AR also provides other feedback that it is recording (Fig. 2b).

By registering for an account users can store audio files; send/receive files to/from other registered users; and, create and enable access to “Groups” of registered users. Users register with their first and second name; a photo, taken using the tablet; and an alphabetic/numeric password, since people use PINs on their phones and many illiterate inhabitants can recognize 10 to 20 names or numbers. Users scroll vertically through profile photos to find their account (Fig. 2a) as inhabitants recognized

people's photos in the media-sharer. To assist identifying files we used a basic folk-somic strategy by displaying those with whom a user shares a file [34]. We symbolized sharing with a palm-up, open hand because it is more relaxed than the deferential local gesture of giving and receiving with both arms.

Users scroll horizontally through profile photos to find other users' accounts to share with (Fig. 2c). Deliberate qualities in speaking and passing messages inspired designing interactions for sharing and 'unsharing' files based on slow gestures not standard ways that aim to increase the speed of interactions [34]. To share a file a user 'long-presses' on a tape icon and drags a small copy of the tape over target users' photos, a collage of Group members' photos, or the public icon (Fig. 2c). The tape drops onto the corner of the photo, when the user lifts his/her finger, where it stays to show the recording is shared.

6 Deploying and Using the Audio Repository

Inhabitants of Mankosi used the AR for 10 months. We introduced and sought opinions about it in 7 workshops lasting 2-hrs to 4-hrs. LRs taught 50 women and men to use the AR, in groups of 2 to 8, and suggested they use Callback [4] to alert others that they had shared a recording. We interviewed 23 people, in four groups, about the AR and owning and sharing digital media and observed useage after deployment. After 9 months we analysed recordings, interviewed users of the AR, and discussed with other inhabitants their perceptions of use and ways to improve access to the AR.

6.1 Use and Potential Use

Inhabitants were very enthusiastic about the AR, and twice 30 women arrived for workshops when we expected six. They said they would use the AR to leave updates about events (e.g. church, choir) or messages for family members that some also said might become mementoes. Community Association members said the AR would help to record meetings as written minutes did not always accurately report what was said. However, despite wide interest, access to the AR was restricted. The Charging Station at Ridge was damaged a week after residents assumed post-trial responsibility for it and conflict arose [7]. The Tribal Authority decided to remove and relocate that Station and tablet and agreed that women could select a new site in Ridge but women did this only recently. Meanwhile, the Headman moved the Station in his homestead away from the office where he stored the tablet and later began to charge the tablet using a system installed for a subsequent WiFi network [6, 35]. He is wary of untrained use and to conserve battery turned on the tablet only to record or listen to files. Thus, people who came to charge their phones did not have easy access to the AR.

The AR was used almost exclusively to record meetings the Headman attended on 17 different days. The Headman stored all recordings in his account: 35 of meetings and 12 that he used to practice or that contained no audible sounds. Mostly there were less 5 recordings a month but in one month there were 17 and there were two months without any recordings. Once the tablet's battery totally drained which meant the date

was incorrect for 50% of files and tablet had been re-set to a Cyrillic script. A third of the files recorded information about issues affecting inhabitants, a third recorded “cases” of tribal law, 20% recorded events in Mankosi (e.g. meetings to end the Stations’ trial) or elsewhere hosted by the Chief or municipality. One file recorded a song from the radio and another a discussion about a useability problem with the tablet. The Headman said the AR was vital as key Tribal Authority members had been less available than usual and said that one day he was alone when another Headman arrived to contest Mankosi’s boundary so, while a child went to alert his secretary, he used the AR to record the exchange. The Tribal Authority said that others were impressed by their advancement in using the AR at Chief’s meetings and five Headmen (governing around 50,000 people in total) had asked how they too could have an AR.

The Headman asked us to delete temporary accounts made in training workshops at Ridge, as other Headmen had reproached him for having too many female friends. Most files (78%) contained only male voices and across the set of files 88% of voices were male. We were able to differentiate between voices in 33 files and found an average of 5.4 different male voices per recording. We discerned an average of 2.9 different females in files containing female voices. There were two files that recorded only female voices amongst a set of five in which a municipality speaker, at the opening of a pre-school, spoke about the use of children’s social grant.

The Headman calls the tablet his “witness” and said he had not deleted any files but he rarely shared recordings using the AR. This might relate to other inhabitants’ comments, in workshops, that there would be “a fight” or “trouble” that entails negotiating with the recorder, if a person contested sharing a recording of them. He recorded topics that he said need to be “referenced” in the future. Some 16% of the files recorded discussions of protocol and inherited authorities or responsibilities and included information about dreams, ancestors and the Royal family. Some of this dialogue occurred in the 14% of recordings that contained disciplining speech acts by, and apologies to, the Tribal Authority. These and other files also related to various disputes within Mankosi or with another community. Most disputes arose in Tribal Authority cases (e.g. paying “damages” to the family of a pregnant woman; fights at soccer matches or within families; physical ostracism and verbal insults). A third of files recorded discussions about environmental resources - 17% explicitly about land, including permission to build; and, 14% about specific assets (e.g. sand). Some 14% of files contained dialogue about the Community Trust, which is expected to manage income from resources and a few files related to local facilities and jobs (e.g. informing workers on local government projects about wage delays).

6.2 Observed Interactions

In workshops illiterate and literate, older and younger, women and men readily learnt and taught each other to use the AR. The Headman eagerly learnt to use the AR (unlike the media-sharer) and taught other Tribal Authority members who learnt less swiftly. Video showed people huddled around tablets, bent their bodies together to listen and held the tablet for each other (Figs. 2b, 2d, 3a). They pointed to the screen to assist others and some made rhythmic hand gestures to depict interactions (Fig. 3b).

The AR orders profiles alphabetically by first name (Fig. 2a) but people found their accounts, or other users, more easily using photos. This is partly because nearly 50% of female first names start with ‘N’ (50% of male names start with ‘S’, ‘M’ or ‘A’). The vertical display of alphabetically ordered photos (Fig. 2a) also affects other meanings. For instance, the Headman said that a photo directly above his own connoted disrespect as it depicted a young man using a rap-style gesture “over” his head.



Fig. 3. Subheadmen (a) and older and younger women (b) gestured (white circle) as they interacted and assisted each other in workshops on the AR. Men carry the tablet (yellow circles) to record community and Tribal Authority members speaking in meetings (c, d).

The Headman decided an issue warranted recording based on who spoke and “what they start to say”. Only two files recorded a whole meeting and files usually recorded one main topic. Some 70% of separate files that were made on the same day recorded discussions of different topics. When separate files were recorded on the same day, at most two files recorded the same or a similar topic. Some 14% of files contained long pauses without speech, this was mostly because in meetings a man carried the tablet between each person as s/he stood to speak (Fig. 3c, d). His movement and the tablet’s size afforded visibility and twice in meetings Community Association members resisted recording as their permission had not been sought.

The tablet’s orientation between people during recording made 9% of files inaudible. Further, 51% of files were recorded outdoors and the wind was audible in a third of these, which made understanding speech in two recordings impossible. Some 21% of files also recorded music, children or other ambient noise.

The Headman used the AR to remind himself of topics after a meeting and “to deliver information” to the Tribal Authority. The Headman listened along with other members, rather than sharing asynchronously, partly because he is the most proficient user. His secretary, however, listened alone to enhance minutes of meetings.

People found a part of a recording easily. Of the Headman's recordings, 22% were shorter than 10-mins; 26% of 10 to 20-mins; 44% of ½ to 1-hr; and, 7% exceeding 1-hr, including one over 4-hrs. He used the time bar that displays when the audio plays to move through the file, but this functionality does not work for recordings of over 1-hr. Inhabitants found a specific file visually less easily, even with few to search. Identifying a recording based on those with whom it was shared did not help the Headman who did not share via the AR. Instead, he looked at the date on the tape, but this was incorrect for 50% of files because the battery had drained.

7 Implications for Design

Use of, and interactions with, the AR shows that portable, asynchronous voice platforms can support oral practices in rural Africa, such as in local governance. To do so, however, it is vital to improve audio quality so it performs better outdoors. Tablets do not yet suit low-electricity conditions but their size and direct-manipulation interfaces offer special affordances for recording and interacting with voice files. Practices in using the AR suggest that accounting for temporal and spatial qualities in co-present interactions might be as important to design as, say, the speed of interacting with information and disseminating it widely. These insights are novel to reports of designing for oral users in technologically sparse settings. Thus, grounding design in local oral practices heightened our sensitivity to conceptual issues that neither cognitive models nor guidelines based on Ong's theory [28] anticipate. We discuss these sensitivities after we outline our current design response to use of the AR in Mankosi.

7.1 Our Voices

Encouraged by enthusiasm for the AR we recently introduced a new prototype for sharing recordings to Mankosi: Our Voices (OV). The Tribal Authority's control of the AR, selectivity in recording and co-present sharing did not widen dissemination but, on the other hand, it improved aspects of information reliability amongst those who have few ways to use phone services. To improve access, and the AR's fragile strategy of manual back-up to external SD cards, OV runs on three tablets at different sites and enables sharing files between tablets when they connect to a wireless mesh network that Jay recently set-up with new collaborators [35]. Community Association members who oversee a tablet in south Mankosi, 1-hr walk from the Headman, say they hope to demonstrate good practice in recording and sharing voice-files using OV.

Although inhabitants appreciated the concept of creating Groups of users the Tribal Authority did not share recordings to Groups in the AR. Nor did they delete files. Further, inhabitants said that a person who records another owns the audio recorded and, if they share it, then all recipients own it. Thus, OV does not include Groups or permit deletion but enables re-sharing files and displays the provenance of sharing.

Many co-present interactions and users' remarks (e.g. the Headman's comparison of the tablet when reset to a Cyrillic script to "riding an unfamiliar horse") shifted our focus to bodily interactions. The AR's deliberately slow interaction for sharing files

[34] resonated with the tempos of physically sharing the tablet in workshops and meetings and suggested that interactions with direct manipulation interfaces can tangibly trace social relations. Thus, OV enables re-ordering users' profiles displayed on each tablet by long pressing and dragging profile photos; for instance, to cluster profiles according to the frequency of using a specific tablet or social roles or relations.

Enabling users to re-order accounts might reduce the time it takes them to find their accounts and other users amongst the wider user base of the distributed OV. Inhabitants do not judge the efficacy of information sharing in relation to Western time scales but electricity is a critical constraint to access, so we included other visual features to speed up search and retrieval. The OV displays the profile of those with whom a user shares most often in a drop-down box. Further, it highlights recordings that users have not yet played and sorts files in the order they were recorded, to avoid the problem of incorrect date-stamps when a tablet's battery drains.

As users sweep through their recordings OV displays a photo reminder of the event recorded by their audio file. Users can take the photo, using the tablet's camera, and attach it to a file at the time or after the event recorded. This aims to assist visually searching for files, but it also offers a way to collect locally created voice and photo instructions to populate a voice-based help and feedback system. Voice-to-text processing for speech recognition demands high quality recordings and computational and electricity resources that are infeasible for tablets in Mankosi. Thus, instead, we are attempting to integrate links between bodies and sounds into search and retrieval. The Headman recognized the topic or setting almost the instant he played a file; thus, as users sweeps through recordings OV starts to play either that file or a user's short vocal abstract about it. Users can record and attach their abstract at the time of recording the main audio file or later, in a similar way as they can tag a file with a photo.

Peer-to-peer synchronization requires temporal overlap when two tablets connect to the WiFi network. This cannot be based on clock-time, since users do not wear watches and the time on electrical devices may be wrong. Thus we are exploring environmental patterns that can prompt switching on tablets to synchronize them, as our ethnography shows inhabitants tune their practices to local rhythms (e.g. sunrise, people or animal movements) [6].

7.2 New Sensitivities in Designing for Oral Users

Developing the AR revealed that advances in audio acutely lag visual technology and our experiences in deploying the media-sharer illustrate that design privileges visual media. This visual bias restricts opportunities to design for oral practices and access by oral users. It contributes to embedding a "language" into interfaces [24] that is further shaped by designers' internalization of print. This design language introduces meanings that can have far reaching ethical consequences [24], even if designers disagree with Ong's evolutionary thesis on the role of writing in 'civilization' [28]. Some who apply Ong's theory recognize that it is the "collective that embodies the shared experience that constitutes the pool of knowledge available to the community" [32]. However, there are no reports of grounding the design of social media in local practices of speaking or situating the design of asynchronous voice systems in

co-present interactions. This contributes to exclusion. For instance, researchers hoped the users they chose to participate in an IVF would transfer information to those excluded from the trial [29] but did not consider how designing the IVF for single-user interactions with phones amplifies existing exclusionary infrastructures.

Our ethnography elsewhere in rural Eastern Cape alerts us to qualities of collective sense-making with recorded information. We (Nic) observed science teaching for 5-hrs in high-school classrooms and every few minutes all the teachers said in English “are we moving together?” or “sihamba” or “sihamba sombani” in isiXhosa, which literally means walking together but also translates to “do you understand?”. They used these expressions most when reading textbooks: they read aloud a piece of text and then, often mid-sentence, paused or asked “and what?” whereupon learners read aloud the end of the sentence or the next. Patterns in linking voices, pauses, meanings and print remind us of Ingold’s (2007) account of early reading. People read manuscripts aloud and treated them as sets of signposts to guide them, Ingold writes, “within the landscape of memory” because meanings *occur* as words are spoken [16].

Nobody experiences reality as a static array of autonomously meaningful things [16] yet we usually start design by reifying events, agents and objects from the relations of practice that give them meaning. Consider, for instance, how modeling names as attributes of a person entity de-contextualises names (e.g. as an attribute “alias” [32]). Mankosi’s inhabitants use various names according to circumstances. For example, they personalize Callbacks [4] with different names depending on why they sent a Callback; and, LRs used different identifiers for the same person in Station logs sometimes combined with other words, such as a prefix like “mama” or “sisi” or the name of one of their children [7]. Names emerge along spatial and temporal paths and recall relations with ancestors through clan, in families by birth order, marriage or parenthood and between peers in shared rites of passage. Importantly, people usually experience names as sounds. Conceptualising words as things that are separate from their vocal, bodily, affective and historical paths omits many relational aspects of sense-making. A heightened sensitivity to this relationality inclines us to embrace the fusion of meanings, sounds and the bodies that produce and experience them.

Contrary to Ong [28], Ingold proposes that it was print, not writing, that reified words because it severed the inscription from the body movement [16]. Interactions with the AR suggest that shared tablets can support the fusion of meanings, the sonority of words and the bodies that speak and hear them to share information. Their direct-manipulation interfaces offer affordances for extra-linguistic qualities of orality; such as those that arise when women in Mankosi sat closely together and gestured as they spoke and guided others to use the AR. The tablet’s portability meant audio files trace spatial patterns such as pauses and sounds recorded along topokinetic paths when a person carried the tablet from one to the next in meetings and indications of the relative orientations of recorders and speakers. We have reported associations between bodies, utterances, spatiality and social relations in sense-making elsewhere in rural Africa [8], and we found that certain meanings emerge in Mankosi when gestures, social roles and the spatial arrangement of photos combined.

Speech sounds and bodily movements are not easily split into discrete events, so we advocate embracing qualities of this continuity in design. For instance, the AR’s

interaction style for sharing reconciles with the temporal textures of movement when people handed or carried the tablet to others or interacted together with it. This is a more nuanced approach to temporality than embedding Western experiences, conceptions and economies of time in design [6], such as to speed up searching because voice recordings listening to them is “tedious” [28]. Guidelines for oral users already suggest that rhythms in verbal utterances aid memory [37]. We propose that responding to other rhythms in information sharing can offer many more affordances in interacting with asynchronous voice, so we end by speculating about this.

In our design workshops for the original digital storytelling application a group of women fluidly took turns speaking when they recorded a story about unfolding events in a child’s morning routine [4]. Their familiarity with each other and their daily routines contributed to this fluidity. The tuning of daily practices to rhythms in the environment may offer a way to prompt inhabitants to switch on and connect tablets to the network to synchronise them [6]. More profoundly, the women’s fluidity in speaking one-by-one and barely without a pause between each affected, and was affected by their interactions with the recording devices [4]. Drawing on diverse research from music to neuroscience, Gill (2012) argues that the rhythmic synchrony of bodies and voices is integral to sociality and enables “being together” [12]. We all have rhythms, in breath, in movement, in gait, that we tune with our speech. Our self-synchrony supports inter-personal coordination by enabling us to adapt our rhythms to others to time our syllables and pauses as we converse. Gill cites a study in which pairs of participants, in isolated rooms who talked using microphones, finely tuned their body movements to their own and the other’s voice. Their rhythms aligned with the language they spoke but concurrent movement also arose without speech [12]. Such synchrony emerges in diverse oral practices in rural Africa - women sing together in harmony, younger people adapt their gestures to elders and protocol, and men respond to each others’ body movements in meetings. This may inspire designing interactions, say to link gestures to vocal input/output, that support sharing information.

8 Conclusion

We designed an Audio Repository (AR) to enable users in rural African villages to record, listen and share voice recordings, for free. Inhabitants learnt to interact with the system easily and illiterate elders, who have few ways to use free or low-cost phone services, used it to record and review community meetings for ten months. We situated the concept and interaction design of the AR in local communication practices and phone use and did not analyze users’ mental functions or apply principles derived from theory about the internalization of writing. Use of, and interactions with, the AR informed the design of a new prototype. Our observations also suggest that the fusion of meaning and sound and tuning of speech and bodily movement can inform strategies for displaying, searching and sharing information. Current design guidelines do not promote such ideas as they focus on single-user interactions and separate mind and body. Thus, we hope to motivate others to enrich the potential of voice-based applications for oral users by situating design in local ways of saying.

Acknowledgements. We are very grateful to all inhabitants of Mankosi, especially LRs. We also thank Thomas Reitmaier for developing the AR and OV, and Simon Robinson and Elina Vartiainen for the media-sharer; Gary Marsden for designing the Charging system; Carlos Rey-Moreno and Zukile Roro for the WiFi mesh; Paula Kotze, Mounia Lalmas and Bill Tucker for support; and, Irina Shklovski, Beki Grinter, Sam Merritt and Heike Winschiers-Theophilus for comments on earlier drafts. Our work was funded mostly by CSIR-Meraka and partially by an UK EPSRC grant (EP/H042857/1).

References

1. Barber, K.: Text and Performance in Africa. *Oral Tradition* 20(2), 264–277 (2005)
2. Best, M., Smyth, T.N., Etherton, J., Wornyo, E.: Uses of Mobile Phones in Post-Conflict Liberia. *Inf. Tech. & Int'l Development* 6(2), 91–108 (2010)
3. Bidwell, N.J.: Anchoring Design to Rural Ways of Doing and Saying. In: Gross, T., Guliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) *INTERACT 2009*. LNCS, vol. 5726, pp. 686–699. Springer, Heidelberg (2009)
4. Bidwell, N.J., Lalmas, M., Marsden, G., Dlutu, B., et al.: Please call ME.N.U.4EVER: Designing for 'Callback' in Rural Africa. In: *IWIPS 2011*, pp. 117–138 (2011)
5. Bidwell, N.J., Reitmaier, T., Marsden, G., Hansen, S.: Designing with mobile digital storytelling in rural Africa. In: *CHI 2010*, pp. 1593–1602. ACM (2010)
6. Bidwell, N.J., Reitmaier, T., Rey-Moreno, C., et al.: Timely Relations in Rural Africa. In: 12th Int'l Conf. on Social Implications of Computers in Developing Countries IFIP (2013)
7. Bidwell, N.J., Siya, M., Marsden, G., et al.: Walking and the Social Life of Solar Charging in Rural Africa. *Transactions on Computer-Human Interaction* (2013)
8. Bidwell, N.J., Winschiers-Theophilus, H., Koch Kapuire, G., Rehm, M.: Pushing Personhood into Place: Situating Media in the Transfer of Rural Knowledge in Africa. *Int'l J. Human-Computer Studies* 69, 618–631 (2011)
9. Blumenstock, E.J., Eagle, N.: Divided we call: Disparities in Access and use of Mobile Phones in Rwanda. *Inf. Tech. & Int'l Development* 8(2), 1–16 (2012)
10. Finnegan, R.: *The Oral and Beyond: Doing Things with Words in Africa*. James Curry/Uni. of Chicago (2007)
11. Frohlich, D., Robinson, S., Eglinton, K., et al.: Creative Cameraphone Use in Rural Developing Regions. In: *MobileHCI*. ACM (2012)
12. Gill, S.P.: Rhythmic Synchrony and Mediated Interaction: towards a framework of rhythm in embodied interaction. *AI & Society* 27, 111–128 (2012)
13. Gorman, T., Rose, E., Yaaqoubi, J., et al.: Adapting usability Testing for Oral, Rural Users. In: *CHI 2011*, pp. 1437–1440. ACM (2011)
14. Harrison, S., Tatar, D., Sengers, P.: Making epistemological trouble: Third-paradigm HCI as successor science. *Interacting with Computers* 23(5), 385–392 (2011)
15. Hull, G., Schultz, K.: Literacy and learning out of school: a review of theory and research. *Review of Educational Research* 71(4), 575–611 (2001)
16. Ingold, T.: *Lines: A brief History*. Routledge (2007)
17. Kaschula, R.: *The Bones of the Ancestors are Shaking: Xhosa Oral Poetry in Context*. Juta, Cape Town (2002)
18. Klemmer, S., Hartmann, B., Takayama, L.: How bodies matter: five themes for interaction design. In: *CHI 2006*, pp. 140–149 (2006)

19. Kumar, A., Sheetal, K., Agarwal, S.K.: Spoken web: using voice as an accessibility tool for disadvantaged people in developing regions. *Accessibility & Computing* 104, 3–11 (2012)
20. Medhi, I., Menon, R., Cuttrell, E., Toyama, K.: Correlation between limited Education and Transfer of Learning. *Information Technologies and Development* 8(2), 51–65
21. Medhi, I., Prasad, A., Toyama, K.: Optimal audio-visual representations for illiterate users of computers. In: 16th Int'l conference on World Wide Web, pp. 873–882. ACM (2007)
22. Medhi, I., Sagar, A., Toyama, K.: Text-Free User Interfaces For Illiterate And Semiliterate Users. *Technologies for Int'l Development* 4(1), 37–50 (2008)
23. Melamed, C.: <http://www.globaldashboard.org/2011/02/07/what-if-you-have-a-mobile-phone-but-you-cant-read/>
24. Merritt, S., Bardzell, S.: Postcolonial language and culture theory for HCI4D. In: Ext. Abstracts CHI 2011, pp. 1675–1680. ACM (2011)
25. Mudliar, P., Donner, J., Thies, W.: Emergent practices around CGNet Swara, voice forum for citizen journalism in rural India. In: ICTD 2012, pp. 159–168. ACM (2012)
26. Murray-Smith, R., Ramsay, A., Garrod, S., et al.: Gait alignment in mobile phone conversations. In: *MobileHCI 2007*. ACM (2007)
27. Ndwe, T., Barnard, E., Mqhele, E., Dlodlo, E.: Effects of application type on the choice of interaction modality in IVR systems. In: South African Institute for Computer Scientists and Information Technologists Conf., SAICSIT 2012, pp. 236–242. ACM (2012)
28. Ong, J.: *Orality and Literacy: Technologizing of the Word*. Routledge, New York (1982)
29. Patel, N., Chittamuru, D., Jain, A., et al.: Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In: CHI 2010, pp. 733–742. ACM (2010)
30. Patel, N., Klemmer, S.R., Parikh, T.S.: An asymmetric communications platform for knowledge sharing with low-end mobile phones. In: UIST 2011, pp. 87–88. ACM (2011)
31. Patnaik, S., Brunskill, E., Thies, W.: Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. In: Int'l Conf. on Information & Communication Technologies & Development, ICTD 2009 (2009)
32. Pitula, K., Dysart-Gale, D., Radhakrishnan, T.: A Case Study in Requirements Engineering for ICT4D. *Inf. Tech. & Int'l Development* 6(1), 78–93 (2010)
33. Porter, G., Hampshire, K., Abane, A., et al.: Youth, Mobility and mobile phones in Africa: a three country study. *Information Tech. Development* 18(2), 144–162 (2012)
34. Reitmaier, T., Bidwell, N.J., Siya, M., et al.: Designing an Asynchronous Oral Repository for Rural African Communities. In: *IST-Africa*. Cunningham & Cunningham (2012)
35. Rey-Moreno, C., Tucker, W.D., Bidwell, N.J., et al.: Experiences, Challenges and Lessons from Rolling Out a Rural WiFi Mesh Network. In: DEV 2013. ACM (2013)
36. Scribner, S., Cole, M.: *The psychology of literacy*, Harvard Uni. Cambridge, MA (1981)
37. Sherwani, J., Ali, N., Rosé, C.: Orality-Grounded HCID: Understanding the Oral User. *Inf. Tech. & Int'l Development* 5(4), 37–50 (2009)
38. Smith, A., Reitsma, L., van de Hoven, E., Kotze, P.: Towards Preserving Indigenous Oral Stories Using Tangible Objects. In: *Culture-Computing*, pp. 86–91 (2011)
39. Sullivan, B.: Freedom to Learn: Blending Interactive Voice Response and Radio. *Learning with Community Media*, pp. 131–135. Commonwealth of Learning (2012)
40. Tacchi, J., Slater, D., Lewis, P.: Evaluating Community Based Media Initiatives: An Ethnographic Action Research Approach. In: *IT4D 2003* (2003)
41. Tucker, W.D., Blake, E.H., Marsden, G., et al.: Reflection on three years of Rural wireless Internet. Protocol communication. In: SATNAC 2007, pp. 452–457 (2007)
42. Walton, M., Hassreiter, S., Marsden, G., Allen, S.: Degrees of Sharing: Proximate Media Sharing and Messaging by Young People in Khayelitsha. In: *MobileHCI 2012*. ACM (2012)