"Huzzah for my thing:"
Evaluating a Pilot of a Mobile Service in Kenya

Jonathan Ledlie

Nokia Research

Cambridge, Mass., USA

Email: jonathan.ledlie@nokia.com

Abstract—Deploying and evaluating a new technology is a major challenge in ICTD research. Introducing new technologies can be hampered by a lack of cultural insight, poor or delayed feedback, and limited evaluation procedures, among other factors. This short paper offers a model for introducing technology in developing regions that mitigates these factors. We call these steps the "Huzzah method," inspired by a quotation that rightly derides technology that is introduced from afar and poorly evaluated. The paper also includes selected portions from other work on Tangaza, whose design, implementation, and analysis followed the Huzzah method.

I. Introduction

In describing the conflicting views on designing and evaluating new ICTD systems, Michael Best succinctly captured how quantitative and qualitative researchers often view each others' work, as observed at ICTD 2009 in Doha:

> An additional tension emerged when those coming from the CS community criticized the social scientific work as lacking rigor or importance. More interestingly . . . was the opposite viewpoint of social scientists finding the work of computer scientists immature. A number of people in Doha described the technology papers to me thusly: "I wanted to build a technology to do this thing. So I started to build it. I did this. Then I did that. Then I did a bit more. Then it was built. Then I asked 10 people from Ghana if they liked my thing. Nine of them did. Huzzah for my thing." [1]

This paper asks: can we as systems builders and qualitative and quantitative researchers really do any better than this? When introducing a new technology, the answer may be no. So what can be changed to turn this imposition of foreign technology and weak evaluation into a sound methodology? I suggest the following approaches, as I will illustrate with our Tangaza project:

- Use local team members
- Act on early feedback
- Acquire honest criticism
- Complement quantitative log data with qualitative surveys

Combining this list with a tongue-in-cheek parsing of the quotation, I suggest calling this methodology *Huzzah*.

When discussing the Huzzah method, I show how its methods were applied in the introduction of Tangaza, a "voice Twitter," in urban Kenya. The appendix includes a series of excerpts from a longer research paper on Tangaza [7]. In particular, the excerpts show how we used local members of our team to design and improve upon Tangaza throughout its piloting phase, and how our mixture of quantitative and qualitative methods complemented each other and provided a richer picture than would have been possible without the use of both.

II. The Huzzah Method

"I wanted to build a technology to do this thing." A danger for any new technology is that the creator simply wants to build it, or to see if it can be built, or to apply some particular algorithm, regardless of the user need. While this *may* succeed in creating a new and useful artifact, it often results in a "hammer in search of a nail," i.e. a problem in search of a solution.

A particular challenge in an ICTD context is when the creator believes some yet-to-be-built technology will be useful, but has limited means to estimate its usefulness without a prototype. That is, walk-thrus, wizards, and surveys can hint at applicability, but nothing can replace having real people actually try the new piece of technology in their real context: the more radical it is, the more people will actually need to try it. Where this becomes fuzzier is with underlying technologies, such as improvements to DTNs (e.g. [5]), where the direct impact on people often cannot be observed within the time frame of the research (or at least not until the end of the

Fig. 1. Interviewing Tangaza trial users from the Huruma Slum in Nairobi, Kenya. We incorporated their feedback into the pilot's later stages.

research). A particularly problematic area here is when the technology is entirely new. However, companies, even small ones, do take this kind of risk quite often.

In building Tangaza, we found one way to mitigate this problem was through having our team include several people from the locale where the technology was (initially) targeted and piloted; in fact, 4/5 of our team are Kenyan. The original idea came from one of the Kenyan members and he and the other Kenyan members helped guide the project toward a culturally-appropriate solution. While the Kenyan members are from a different socioeconomic class from Tangaza's target group of lower income users, these team members were a source of constant insight; we did not fret over imposing a culturally-inappropriate technology for this reason. Thus, when "building your thing," it is extremely helpful to have many if not most team members come from the place where it will be deployed. In cases where this is not possible (and it probably is), partnering with a local university or company may be the best alternative.

"So I started to build it." As with any project, this step must include an analysis of prior work: have others (not just academics) built anything similar in a similar context? Or are there qualitative analyses that can be used to make certain design decisions?

An additional step that we found particularly useful in Tangaza's development was user interviews performed directly by the team of designers and implementers. While the initial presentation of the technology occurred at our lab, the interviews were in one of the trial participant's homes, near where most of the participants lived. The users themselves made several suggestions for new features which we could debate right then and have implemented and part of the live pilot a few days later. Because members of our team spoke the same local languages and overlapped in ages with many of the participants, it was easy to have unstructured and open early feedback. In addition, many participants appeared to find this process empowering. This early feedback helped us improve the system at a much faster rate than if people had only been surveyed at the end of the three month trial.

A related problem is balancing (a) acquiring feedback from end-users against (b) biasing their eventual opinion of the technology. By showing and discussing early prototypes, users may not understand the technology as it has become, rather as it was when they first understood what it might do. This can be avoided with multiple focus groups, but interviewing disjoint sets of people may not be possible in certain projects or with a limited time budget.

An alternative approach to rapidly incorporating feedback into a prototype during the development process is to prepare a brief interview for early users immediately after they have completed a task. For a different project that is in earlier stages, we plan to automatically call a subset of users right after they have used the system, and connect them to a member of our research team who speaks several local languages. With this approach, users can be given a quick open-ended survey and, because the call will be with a real person, the questions can veer into a dialogue if the user wishes.

"I did this. Then I did that. Then I did a bit more. Then it was built." Before one declares the system "built," one needs to ensure that the channels for feedback from participants are valid and constructive. In many cultures, people are predisposed to say yes; they do not say what they think, but what they think others want to hear.

It was surprisingly not difficult to elicit honest criticism from our trial participants, even though the lower-income group was paid. Two key actions resulted in clear feedback from the Huruma Slum group (Slumcode) in particular. First was the way in which the project was initially presented to the group: as an early trial of a new technology where their input will help guide its future. One of the Kenyan team members gave an overview of the technology (in English and Sheng); our team lead spoke only briefly at the beginning and end of the session, where the central emphasis was on the type of feedback we would like to receive ("Please be critical!") and not on the technology itself. Second was the informal meeting with the Slumcode group at one of their homes (Figure 1), where the head of the Slumcode group set the tone by starting off with constructive, but significant criticism. Following this, other group members were very forthcoming with positive and (more importantly) negative feedback. Again, having most of our team be able to speak in the local languages was also extremely helpful and made it simpler to speak candidly.

"Then I asked 10 people from Ghana if they liked my thing." Well, at least these hypothetical authors *asked* people instead of simply parsing log files. This is the core issue: how was the "thing" evaluated? Which people were asked and how, and what were they asked? Did the qualitative results from surveys and questionnaires match quantitative results or contradict them – why or why not? Presumably this agreement will occur in some parts of the evaluation and not in others. How did the quantitative and qualitative results differ from the researcher-designers' expectations (because you, the evaluator, must have had certain expectations, or else you wouldn't have built your thing the way you did).

Because Tangaza runs on a server, we had very detailed logs of *what* users were doing. However, we had only limited information on *why* they performed certain actions. One (Kenyan) member of our research group is a sociologist; she guided the creation of a detailed survey, which included both multiple choice and open-ended questions. Because people listed their phone numbers on the survey, we were able to then correlate specific users with quantitative results from the logs. This matching of self-reports to actual usage was critical to gaining a full picture of how Tangaza was being used. For example, we could see when a user who described Tangaza as "expensive" stopped using it, and how others who stated they preferred voice tangazos over group texts usually did, in fact, send more voice tangazos.

In addition, the qualitative and quantitative data often covered distinct topics. We could not have known about

people's attitudes about security and privacy without asking them (§D3), and observing the change in group structure over time (Figure 5) or the length of tangazos (Figure 4) would have been a challenge without logs.

A challenge here is thinking ahead to the quantitative questions you, the evaluator, will want to ask at the end of the trial, and making sure the logs sufficiently record the necessary data. In addition, you must be prepared for the time and expense of not just asking nine people in Ghana, but a substantial diverse slice of your users (we could have benefited from having more than thirty participants complete our survey).

"Nine of them did. Huzzah for my thing." Building, deploying, and evaluating a new project, particularly in a developing country, is typically not a clean and structured process. Often one must make an educated, locally-informed guess and then just build the thing. But by actively seeking out criticism, and by rapidly altering your prototype, and by cross-validating qualitative and quantitative evidence (and by being willing to fail), you may well be able to say "Huzzah for your thing."

APPENDIX

The appendix is a series of excerpts from a longer research paper on Tangaza [7]. I have included a background overview section (§A), design process (§B), evaluation methodology (§C), and a few results to highlight the qualitative and quantitative methods used (§D).

A. System Overview

Tangaza is a social networking service that makes use of SMS and voice to enable low-cost group communication. Users can create and manage their groups and invite others to join. After a group is formed, members can then text and send spoken updates to each other. Through sets of SMS-based commands users can create both public and private groups, giving them the flexibility to decide who can or cannot be members. After a message is sent, the members of the group are notified about the existence of a new message via a "flash," i.e. missed call, or a text.

To better understand Tangaza's design and implementation, we first give an overview of how users create and join groups and send each other spoken and text messages with only a basic mobile phone as a requirement. Figure 2 shows the process of a person A creating a group and then inviting two other people B and C. After they accept the invitation, A sends the group a text message, which is delivered to both B and C. Illustrating a common use of Tangaza, the message sets up a meeting time for group members. After the meeting, person B sends a spoken message, or *tangazo*, to the group. Tangaza flashes A and C, notifying them of the new tangazo, which they call in and listen to.

While the example shows the people as having called Tangaza directly – bearing the expense themselves – Tangaza limits end-user costs and increases usage by allowing users to flash it, which results in a call-back from the interactive voice

Fig. 2.   Tangazo and Text Example

Create	A send:	`create testgroup 2`
	A recv:	`OK. Created the public group testgroup, assigned key 2.`
Join	A send:	`invite testgroup 0722981234 0719115678`
	B,C recv:	`A invited you to the testgroup group. Reply: join testgroup.`
	B,C send:	`join testgroup`
	A recv:	`A new user B<0722981234> has joined testgroup`
Text	A send:	`@testgroup Meeting is at 6pm tomorrow`
	B,C recv:	`Meeting is at 6pm tomorrow A@testgroup`
Tangazo	B calls:	*Selects key 2*, *Records* "Poa, was great to see you both!"
	A,C alerted via flash:	*Each observes existence of new tangazo*
	A,C call:	*Each listens to B's message*

response system (IVR). This mimics a common mechanism in developing regions in which people regularly communicate through missed calls [4]. We use only the most common meaning of a missed call, or "flash": call me back.

As the example shows, Tangaza is used exclusively through texting and calling: *no* software is installed on the user's phone. We discuss how we came to this decision, its advantages, and other design choices in Section B.

B. Design Process

Our internal design and prototyping of Tangaza involved several iterations. In particular, our initial, internal designs did not use text messages at all. Instead, both group management and tangazos were conducted through an IVR; some initial designs used speech recognition, some touch-tone-based input, and some a combination. For example, a user could pre-record a name for a group and then use voice to look up this name during a later call. This combination, however, had complicated menus and proved cumbersome and error-prone. As a result, the system was split into SMS and IVR components, with group management tasks ("control" tasks) shifting to SMS and with listening and sending of spoken updates as the sole tasks for the IVR. We also removed speech recognition entirely from the IVR because, as others have shown, the resulting ambiguities and mistakes tended to frustrate users, particularly in noisy environments [3], [8]. Individual group naming was also confusing because people often supplied different names for the same group.

After releasing Tangaza externally to our test users, we continued to refine the user interface based on their active feedback, relying on Participatory Design techniques [2]. Through

TABLE I
POPULATION DEMOGRAPHICS

	Strathmore	Slumcode
Income	medium - high	low
Participants	87	13
Age	20 - 26	18 - 40
Sample occupations	students	caterer, social worker, unemployed, artist
Education	University	High school, College
Min/Call	5	2
Calls/Day	3	3
Texts/Day	5	3
Budget/Day	137.5 Ksh	58.5 Ksh

videotaped interviews with groups of participants early in trials and through on-going contact with the groups [10], we learned about several aspects that users found confusing and about their privacy concerns. We also received active confirmation that a primarily spoken-message group communication mechanism like Tangaza would be useful to them socially and commercially. Several members of our group speak Swahili and Sheng, a local English and Swahili patois; we communicated in these languages during the feedback process. Figure 1 shows a photo from these interviews.

We drew early feedback mainly from technologically-savvy "key informants" from both our low and moderate-income populations [6]. In response, we refined the invite/join protocol to be more informative at the expense of more messages (and therefore less frugal). In our original version, no message was sent to the inviter when the invitee accepted an invitation. Users told us that they assumed they would be notified and, when they were not, they assumed a system or network error. We changed this so that the inviter was notified by SMS whenever the invitee confirmed joining the group. In addition, we added the ability for users to give themselves a pseudonym instead of having their phone number displayed on each message. As we discuss later, this only partially ameliorated security and privacy concerns.

C. Evaluation

In order to learn about how different groups of people in a developing country would understand and use Tangaza, we deployed it in Nairobi, Kenya. We recruited two different user populations to try it, invite their friends and family, and give us feedback. We used this qualitative feedback along with quantitative data from logs to examine: how the two populations differed in their perception and use, how groups formed over time, and how they perceived the user interface.

1) Two Pilots: We deployed two separate pilots of Tangaza from February to April 2010 in: (a) Huruma Slum and (b) Strathmore University. Table I summarizes the demographic differences between the pilots' populations (Phone data are per-person medians of previous day's reported usage; budget estimate assumes in-network calls and SMS with a prepaid account).

Slumcode (Huruma Slum). Our user population representing low-income users in a developing country came from a youth group called Slumcode. Slumcode focuses on personal development, primarily through community-based activities and technology education, and consists of twenty young people. The group is based in the Huruma Slum, a dense, low income settlement in Nairobi. Slumcode was selected because of its previous work with our organization; all members were invited to join, and thirteen people became active Tangaza users. Slumcode members have varying degrees of technological exposure and many have extended rural family members who have limited exposure and low literacy levels. In addition to free use of the service, the group was paid for its participation in the trial.

Strathmore University. Our second user population came from Strathmore University, a business and IT-focused college in Nairobi. Strathmore students primarily come from middle and upper income levels; they are technologically adept, and active, but not always-connected, Internet users [12]. Out of similar university settings, we selected Strathmore because a member of our research team was a recent alumnus and able to orchestrate a high level of interest from students and faculty. While over 440 people from Strathmore have used Tangaza, we focus on 87 active users during the pilot period. While their use of Tangaza was reimbursed, the participants themselves were volunteers.

2) Methodology: We isolated the log records of active Tangaza participants during the trial period and asked them to complete an evaluation and impact assessment survey. Eleven Slumcode participants and nineteen Strathmore participants completed the survey. Because the Strathmore students were on holiday, they were given the option to complete the survey online. Participants were paid a small amount for completing the survey.

The survey examined comprehension, usage, likes, dislikes, preferences, impact on communication, ease of use, related costs, and comparison to other communication systems. In addition to the survey, the system logs showed how, when, and how often participants used Tangaza. By examining a subset of the texts and tangazos, we also gained some insight into the purpose of different messages and groups.

D. Results

1) Comparing User Populations: The two pilot groups used Tangaza in ways that primarily reflected differing levels of cost consciousness.

Aggregate Usage Patterns. We first examine aggregate usage of the pilots, as shown in Figure 3. From the data, we can make three observations. First, the Slumcode users texted far more often than Strathmore ones. This reflected a commonly stated preference among Strathmore users for tangazos over texts. For Slumcode users, this appeared to be an unconscious continuation of their standard behavior: because texting is a significantly cheaper method of communication outside of Tangaza, these users continued to text. Outside of Tangaza, Slumcode users allocate 50% more of their phone budget to texts than Strathmore users, at 18% vs. 12%, respectively (they text less overall because their total budgets are less
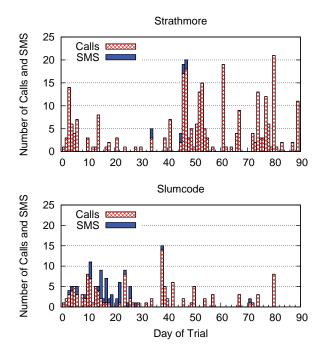
Fig. 3.   Aggregate Calls and SMS during trial

Fig. 4.   Length of tangazos exemplify the greater cost consciousness of Slumcode participants.

than half of Strathmore users). Slumcode users also expressed how tangazos could be difficult to send or receive in noisy environments, such as buses, which are more common for this group.

Second, the per-person usage of Tangaza was significantly greater for the Slumcode group. While Strathmore had more calls in total due to their larger population, Slumcode had on average 65% more calls and 80% more texts per person per day. While this is due in part to a more focused pilot, it also reflects a stronger relevance of the service, as compared to other social networks that the Strathmore group had easy access to. Note that the SMS data only include sent group texts, not commands (e.g. `join`).

Third, Slumcode use diminished during the second half of the pilot. This occurred because a service provider did not have automatic reimbursements properly implemented at the beginning of the trial and, later, failed to note when the reimbursements did not succeed. Among this group of users – who knew their exact mobile credit – this led to significant frustration. While some users resumed using Tangaza, many remained distrustful, with some calling the service "very expensive."

Cost Sensitivity. The population demographic data in Table I showed that Slumcode users tended to have short, frequent calls. Because mobile billing is per-second in Kenya, low-income people have even evolved a short-hand speaking notation to save money: calls tend to be short and to the point; this is in addition to using missed calls to transfer or eliminate calling expense.

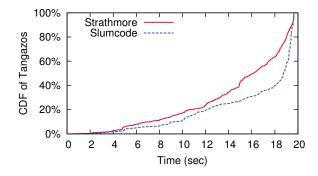We found this maximization of resources extended into their use of the Tangaza pilot, even though the pilot was free. Figure 4 illustrates this pattern in comparison to the Strathmore participants, who do not appear to be as cost-conscious. Slumcode users consistently used almost all of their allotted twenty seconds per tangazo. In discussing this limit, two Slumcode users expressed a desire to leave multi-minute tangazos, but most appreciated the known cost that short tangazos would bring for both senders and receivers.

While Slumcode users are more cost-conscious, both groups wished to limit expenses: 87% of all participants said they would prefer a prepaid tangazo/text allocation plan, which would prevent unforeseen high costs.

2) Group Formation and Evolution: Growth and usage during the Tangaza pilots mirrored typical social participation rates, such as the Pareto contribution rates in Wikipedia.
Why some groups grew. The most successful groups were based on pre-existing, non-virtual connections. For example, the "slumcode" group was active, as were several class, homework, and club groups within Strathmore. Slumcode also included groups that were used for church and family activities. Active groups tended to have a small subset of active members; for example, while everyone in the Slumcode pilot was a member of the group "slumcode," there were only four highly active members.

Figure 5 shows the change in group membership over the trial period for both pilots. It illustrates how the Slumcode participants, in particular, created many groups which often went unused; they did have three active groups, and one small group with an external member (10). Strathmore users tended to form larger groups based on class and club memberships. Several smaller groups were used for homework and socializing. Within groups, texts were primarily used casually and for greeting, while tangazos were for more serious notifications such as changes in meeting schedule, homework discussions, and coordinating events.
Why others did not. While the more active groups were based on pre-existing local networks, some groups became dormant, mainly because the scale of the Tangaza pilot did not match the ubiquity of other online social networks. This was particularly true for the Strathmore group. Strathmore users said Tangaza was "not widely used yet," that "more people need to know of its existence," and "other people stopped *tangazaring* so I
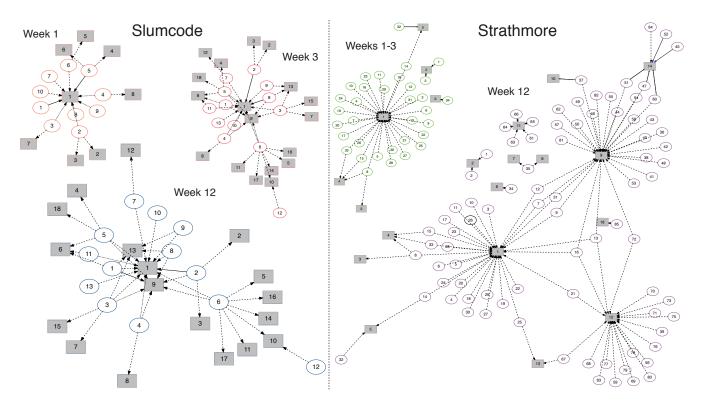
Fig. 5. The graphs illustrate how the constituents of the Slumcode and Strathmore groups changed over the trial period. Each box represents a group; each circle represents a person. Less active group membership is shown with dashed lines; solid lines show active membership.

also followed suit."

3) User Interface: Our design attempted to make creating and listening to new tangazos fast and, therefore, cheap. While most people said Tangaza was "somewhat easy" to use, some tasks were harder than others. People from both groups felt that sending a message was particularly easy, but creating groups and inviting people to them was particularly difficult. Many people wanted to use invalid characters (punctuation and spaces) in group names – e.g. "Deno R.I.P.," "odhoji's" – and were frustrated when these were rejected. Perhaps not surprisingly, some people expressed that there was a disconnect between which key needed to be pressed to select a group and the group itself. Users also wanted reinforcement that they had selected the correct group when using the IVR; we had not provided this to speed up the interaction. Instead, announcing the name of the selected group, as recorded by the group creator, could provide this feedback. Another option would be to use speech recognition for group selection, where either the user or the administrator would have supplied the matching utterance; recognition and feedback in low-resource languages can be particularly tricky, however [9], [11].

Privacy and Access Restrictions. Several Slumcode participants and one from Strathmore used the nickname feature to hide their identities. Many in Slumcode, in particular, continued to find privacy a concern, stating that "it shares the number with all the numbers in the group." This was due to not knowing about the feature rather than a problem with the feature itself: fewer than half of the Slumcode participants remembered about nicknames whereas 2/3 of Strathmore knew about this feature.

All groups used the default "public" level of access. Because participants knew each other, it is unclear if our proposed access controls would, in fact, map well on to larger groups, which would presumably be more of a management challenge.

REFERENCES

[1] M. Best. Workshop on Computer Science and Global Development. Aug. 2009.
[2] A. Dearden and H. Rizvi. Participatory IT Design and Participatory Development: a Comparative Review. In *PDC*, Oct. 2008.
[3] C. Delogu, A. D. Carlo, P. Rotundi, and D. Sartori. Usability Evaluation of IVR Systems with DTMF and ASR. In *ICSLP*, Nov. 1998.
[4] J. Donner. The Rules of Beeping: Exchanging Messages Via Intentional "Missed Calls" on Mobile Phones. *JCMC*, 13(1), 2007.
[5] S. Guo, M. H. Falaki, E. A. Oliver, S. U. Rahman, A. Seth, M. A. Zaharia, U. Ismail, and S. Keshav. Design and Implementation of the KioskNet System. In *ICTD*, Dec. 2007.
[6] D. R. Millen. Rapid Ethnography: Time Deepening Strategies for HCI Field Research. In *DIS*, Aug. 2000.
[7] B. Odero, B. Omwenga, M. Masita-Mwangi, P. Githinji, and J. Ledlie. Tangaza: Frugal Group Messaging through Speech and Text. Dec. 2010.
[8] N. Patel et al. A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. In *CHI*, Apr. 2009.
[9] M. Plauche et al. Speech Interfaces for Equitable Access to Information Technology. *ITID*, 4(1), 2007.
[10] Y. Schwartzman and T. S. Parikh. Establishing Relationships for Designing Rural Information Systems. In *CHI*, Apr. 2007.
[11] F. Weber, K. Bali, R. Rosenfeld, and K. Toyama. Unexplored Directions in Spoken Language Technology for Development. In *SLT*, Goa, India, Dec. 2008.
[12] S. Wyche, T. N. Smyth, M. Chetty, P. M. Aoki, and R. E. Grinter. Deliberate Interactions: Characterizing Technology Use in Nairobi, Kenya. In *CHI*, Apr. 2010.