# Hyke: A Low-cost Remote Attendance Tracking System for Developing Regions

Azarias Reda
University of Michigan
azarias@umich.edu

Saurabh Panjwani
Microsoft Research
saurabh.panjwani@microsoft.com

Edward Cutrell
Microsoft Research
cutrell@microsoft.com

## ABSTRACT

Tracking attendance is an important consideration for many developing world interventions. In many cases, these interventions are located in remote areas where its not always feasible to deploy expensive attendance tracking systems. In addition, since many existing systems focus on tracking participants (such as patients or students), rather than agents (such as teachers or health workers), they assume a trusted administrative staff on-site to record attendance. In this paper, we present the design of Hyke, a system for remote and cost effective attendance tracking in developing regions. Hyke combines voice-biometrics with accurate location tagging for tracking attendance in remote locations without the need for a trusted mediator on-site. Hyke was designed based on our observation of a currently deployed teacher attendance tracking system in rural Rajasthan, India. We have implemented some of the key components in Hyke, and discuss some of the security concerns in the system. The Hyke biometric stack for voice recognition is built atop several open source technologies, and provides a simple interface for non-expert users. Our evaluations with Indian speakers over telephone audio suggests the biometric stack is at par with the current state of the art. We believe this will be a useful tool for researchers who would like to incorporate voice technologies in their developing world projects.

## Categories and Subject Descriptors

D.2.0 [**Software**]: Software Engineering

## General Terms

Design, Management

## Keywords

attendance, voice, challenged networks

## 1. INTRODUCTION

Tracking attendance of remote agents is a fundamental problem that several developing-world applications require for sustained operation and impact. Healthcare outreach programs must monitor the arrival of health workers in designated locations and at designated times in order to ensure benefits to the intended beneficiaries. Education-focused institutions must stay aware of teachers' presence or absence in respective classrooms to ensure benefits to students. In several such applications, the agents who need to be tracked operate from remote locations which makes monitoring their attendance a non-trivial task.

Numerous studies report a high rate of delinquency of service personnel in healthcare and education programs in developing regions [6, 10]. Recent interventions have tried to make use of electronic attendance-tracking technology to address this problem. One pioneering effort in this direction was taken by a non-profit named Seva Mandir in southern Rajasthan, India [1]. Seva Mandir uses photographic information from digital cameras to track attendance of teachers and health workers working in rural areas, and provides monetary incentives to increase attendance. More recently, three government bodies in India (in Gujarat, Haryana and Noida) have announced the deployment of fingerprint scanners to track attendance of teachers, with one pilot underway [2]. Others have attempted to track the attendance of patients in health programs using fingerprints [15].

While current electronic attendance tracking systems provide several benefits over manual approaches, most of them are still too expensive for large-scale deployments. In addition, many existing systems require a trusted mediator on site, which can be difficult when tracking the attendance of remote agents. To address these issues, we present the design of a new system, Hyke, specifically for low-cost remote attendance tracking in developing regions. Hyke uses mobile phones as the platform of authentication and builds on open-source voice biometrics technology in combination with off-the-shelf location tagging tools.

Hyke has several advantages over prior systems. Besides reducing cost of operation, it offers the possibility of doing attendance tracking without the presence of a trusted administrative staff on-site—both location and timestamp information for attendance records are generated automatically. Another advantage lies in its utilization of voice as a user biometric—voice is generally regarded as a less invasive and privacy-sensitive form of biometric than fingerprints or pictures. Hyke uses widely available cellular networks with voice and SMS channels for communication.

This paper presents an initial design description of Hyke, including a prototype implementation of some of its key pieces. Perhaps the most critical component in deploying Hyke is its voice biometric stack. An important consideration in building the stack was providing a simple interface for

*NSDR'11,* June 28, 2011, Bethesda, Maryland, USA.

a voice biometric toolkit that can be reused by other projects for developing regions. While voice recognition technologies have been previously tested in several contexts [8, 13], existing tools are generally not targeted to the non-expert researcher. Our biometric stack builds on several open source tools, and provides a simple interface to an end-to-end solution at par with the current state of the art. Our experiments with Indian speakers using audio collected over telephone shows error rates less than 5%, providing sufficient accuracy for most applications.

We start out by describing a field visit to Seva Mandir in Rajasthan, India, which formed much of the motivation for our work. In section 3, we present the design of our system and provide more details in sections 4 and 5. Related work and conclusions appear in sections 6 and 7.

## 2. CASE STUDY: SEVA MANDIR

The design of Hyke was motivated by our visit to Seva Mandir's field operations in November 2010. Seva Mandir currently runs a set of 200 non-formal primary education centers in rural areas of southern Rajasthan. Each "school" employs 1-3 teachers, and serves 20-60 students in grades 1 though 5, and is located up to 150km from the NGO's headquarters in Udaipur. In response to rising concerns of teacher absences, the NGO has been using camera-based attendance tracking system for teachers [10]. Using commodity digital cameras provided to each school, teachers are required to take their photographs in the school setting twice a day. These photographs, along with the timestamps, are used as evidence for attendance, which is later used to compute monetary incentives.

In our visit, we studied the implementation of camera-based tracking in 7 schools, and conducted interviews with 10 teacher and 12 staff members, along with informal discussions with over 20 parents and over 100 students. Somewhat surprisingly, a majority of the teachers expressed optimism towards the intervention, highlighting how it had made them more disciplined and responsible towards the issue of attendance. Students and their parents consistently reported in favor of the system and students seemed excited by the opportunity of being photographed that it offered.

While positively perceived by several users, there were some drawbacks of the system that surfaced. By far, the biggest drawback was the human-intensive process of picture verification—the system currently relies on 4 employees of Seva Mandir who physically verify photographic data from all schools (a total of nearly 18,000 pictures per month), and compile it into a list of attendance sheets. The entire process requires roughly 100 man hours on a monthly basis (equivalent to $100-200), a significant investment in the context of developing regions. Another limitation was the manual process of ferrying memory cards from the schools to the verification centers at the end of each month, which caused delays in attendance processing, payment and dispute resolution. Even though more than 75% of the schools were under cell-phone coverage, the cost of using data networks and the need for more devices at the schools has prohibited digital transfers. Furthermore, since memory cards are not suitably protected in the current system, the timestamp and picture information they carry are susceptible to forgery—EXIF data on picture files is easy to modify for a sufficiently-determined user (although we did not observe any incidence of such an attack).

## 3. SYSTEM DESIGN

Any electronic system for tracking attendance of remote agents must make three types of measurements: it must record the *identity* of the agents, it must match their *location* with that of a given facility and it must note the *time* the agent is located at that facility. All three variables are equally important and errors in measuring even one affects overall system functionality. It is also important that the identity of the agent be verifiable in an accurate manner by the system and that location and timestamp information cannot be spoofed. Finally, from the point of view of developing regions, it is essential to minimize costs and maximize usability by novice technology users.

Hyke is designed with these goals in mind. It leverages voice biometrics for identity verification of users and couples it with a location tagging mechanism and a network-based time-stamping protocol. It is designed to work in environments where it is difficult to ensure a trusted administrator on-site, and connectivity is largely limited to cellular networks with voice and SMS. This covers a large number of use cases in developing countries, and the Seva Mandir school system is a good example. Hyke provides an automated alternative for attendance tracking in these environments.

Hyke consists of three main components—location tagging, passcode generation and biometric verification. We will begin with a higher level description of the system, and then discuss each component. We will also mention some of the security implications for our design decisions, and return to some of them in section 4.

During setup, the facilities where attendance is required (e.g., schools, health centers) are tagged and associated with a unique ID. Each agent whose attendance is being tracked submits voice samples which are used for identification later on. A central verification server records IDs of all facilities and voice samples of the agents. Each agent is equipped with a Java phone that runs the Hyke client. The client identifies its location either through a GPS antenna, or using an RFID tag that is in its vicinity. When an agent is ready to record her attendance with the system, she starts the Hyke client which obtains the location from the location tagging component and sends it to the server via SMS.

The server responds with a one-time passcode, which is then displayed to the agent on the phone's display. Afterwards, the client also dials the Hyke attendance hotline. Upon connection, the user simply provides her ID and then reads out the passcode to the system. On the backend, Hyke verifies the user though speaker voice identification, and extracts the passcode using simple speech recognition. The system uses this information to verify the claimed user and location on the backend, and record attendance. Hyke includes several mechanisms designed to prevent record/replay and man-in-the-middle attacks, as well as a secondary failover strategy for gracefully handling uncertainties in biometric verification.

### 3.1 Location tagging

Hyke allows two mechanisms for location tagging—GPS or a physically-secured RFID tag. Computing a location ID from GPS coordinates is fairly straightforward, and can be accomplished using predefined grids. In RFID-based location tagging, a passive RFID tag with a unique ID is physically secured to the facility and the agent is required to bring her phone in the vicinity of the tag for the Hyke client to read the tag's ID. RFID tagging might be useful when dealing with locations such as office buildings where it's difficult to get a GPS signal. One way to secure an RFID tag might be to print the tag on perforated paper and glue it to a concrete wall, so that an attempt to remove the tag will destroy it.

We remark that fuzzy location identification mechanisms such as radio triangulation are not appropriate for remote attendance tracking due to the strict location requirements in this application and the lack of sufficient radio beacons in remote locations. A location reading in Hyke is never cached on the client phone, and is freshly obtained every time attendance recording is initiated. This lets the system know the phone was in the target facility at the time of reading. This will later be extended to verify the agent was at the location. GPS-enabled Nokia phones are now available for less than $50, while RFID-capable phones are approaching the $100 price-point.

## 3.2 Passcode generation

Hyke uses one time passcodes to guarantee freshness of the users voice. Once the Hyke client has read the location ID, it sends a hello SMS message to the server containing the ID. The server uses this SMS to verify the location reading phone is the one associated with the location in the Hyke database. This can easily be done by observing the 'from' field in the message. If this is the case, the server generates the one time passcode that's sent back to the phone, as well as stored on the server with an expiration time. A passcode can only be used within the specified time to guarantee freshness and at any time, there is at most one valid mask for a client phone. If either of the SMS messages fails to be delivered, no code will be displayed, and the agent is notified to try to read the location ID again. Once the passcode has been displayed to the agent, she is expected to make a note of it for the verification step. To simplify implementation, Hyke uses numeric digits for a passcode. As an added benefit, limiting the vocabulary requirements for the speech recognizer allows for wide support of local languages [16].

There are some obvious network security threats to this exchange between the client and server. SMS's can be eavesdropped upon (using special-purpose hardware) and the sender ID of an SMS can be spoofed. Such attacks are possible to prevent using standard cryptographic tools and we omit details here for brevity. Our expectation is that, in practice, the gap between the educational background of our target agents and the technical sophistication required to carry out such attacks will be significant enough to reduce attack likelihood.

## 3.3 Biometric verification

The third component of Hyke's design is biometric verification of users. Hyke leverages voice for doing remote attendance verification. This enables organizations to deploy Hyke in areas where its not feasible to have computer based attendance systems. In addition, the wide availability of cellular coverage in developing countries makes it a low-barrier alternative. While we considered image recognition based verification initially, the difficulty of performing the computation on resource constrained mobile phones, and the network cost of shipping images to a remote server strongly favored voice.

Speaker recognition technologies have been steadily improving over the past decade. The National Institute of Standards and Technology(NIST) conducts annual competitions that gauge the current state of the art in speaker recognition, and the results have been getting increasingly better [13]. Speaker recognition requires users to enroll in the system. This is used to train a representative model of a user's voice that is used to compare against future utterances. Once a model is built to represent a user, speaker recognition can be used for verification or identification. Speaker verification is a 1-to-1 comparison between two models to determine if they are from the same person. Speaker identification is a 1-to-N comparison to identify a speaker from a set of N candidates.

There are two components to Hyke's biometric stack. The first one does speaker verification for recognizing users, and the other does speech recognition for extracting digits from the one time passcode. Speech recognition is a more widely studied problem, and several out-of-the-box solutions exist for it, including programming frameworks in Microsoft Windows and Mac OS. Therefore, we will focus more on speaker recognition in this paper.

When a user calls the Hyke hotline, she is requested for her unique ID. Afterwards, she is prompted to read the one time passcode displayed by the Hyke client. As she utters the passcode, Hyke first does speaker identification to verify this user is indeed the one identified by the user supplied unique ID. Simultaneously, Hyke uses digit recognition to extract the passcode from the utterance. This passcode is used to verify the freshness of the attendance report.

Our experiments using local Indian speakers with telephone recorded voices shows Hyke's speaker verification results are at par with the current state of the art, with over 95% accuracy rate on average. Since we imagine a speaker recognition toolkit could be useful in many developing region projects, the Hyke biometric stack is packaged separately. Our key contribution in this regard is providing a simple interface for an end-to-end speaker verification systems that can be used by non-experts. This will allow other researchers to easily incorporate speaker recognition technologies in their systems. The speaker toolkit, as well as data collected from local speakers over the phone, will be publicly available.

## 3.4 Fail-over mechanisms

Even with over 95% accuracy rate, its still important to gracefully handle uncertainties in biometric verification. If the system can not confidently establish the identity of a user after a threshold number of tries, Hyke invokes its fail-over mechanism. In the case of Seva Mandir, this mechanism is leveraging the digital cameras already available in the school. If a teacher's voice can not be identified for any reason, the attendance hotline informs her to take a picture instead. In practice, the fail-over mechanism can also be invoked randomly to increase the system's robustness against forgery, or when the system cannot be used due to network coverage issues.

A more elegant approach is to redirect callers who can not be automatically identified to live operators. Since teachers already call the district office for attendance if there is any problem with their cameras, we will be leveraging an existing framework for verifying uncertain cases. A live operator can verify attendance through a number of alternatives, including personally identifying questions or leveraging students present in the school. Appropriate fail over mechanisms are going to be different for different organizations, and Hyke can be customized to support them.

## 3.5 Attendance presentation

The final step in completing the attendance tracking loop is presenting the attendance report to the organization. Hyke can provide a CSV or XML formatted report that can be imported to existing programs for visualization or integration with other data. While Hyke focuses on the attendance tracking process itself, it could be useful to build a web portal that displays attendance in a more visually pleasing manner. It is generally straightforward to take a Hyke report in a standard format, and build a custom visualization portal for attendance.

# 4. SECURITY

In this section, we will discuss a few more security threats to the Hyke attendance system. While most common attacks such as eavesdropping and impersonation can be thwarted with standard cryptography, others are more specific to the Hyke system. We will focus on the later.

## 4.1 Conference calls

One mechanism to beat a voice based attendance system is for an accomplice to initiate a conference call from the attendance phone, such that the attacker's voice would appear to come from the target location. As the server identifies the calling phone as that of the attendance phone, the attacker then would be able to forge attendance. We have two mechanism to prevent this. The first is disabling the conference calling feature from attendance phone's service by contacting the cell phone provider. The second is to initiate calls to the hotline programmatically, and disable user-input when the Hyke client is running.

## 4.2 RFID cloning

When using an RFID based location tagging, it might be important to prevent cloning. While such attacks might not be practical in our use environment, encrypting RFID signals can provide a sufficient protection against such attacks. An RFID chip capable of encryption costs about $5 compared to a typical passive RFID tag that costs under a dollar.

## 4.3 Reverse engineering

To verify attendance location, the Hyke client relies on its reading from the location tagging component. If an attacker is able to reverse engineer the client so that this location information can be provided by a custom component that always provides the correct location, the system breaks down. Once again, this might not be practical in our use cases due to the advanced technical skills needed, and obfuscation and binary encryption tools for modern programming languages. However, Hyke attendance phones are also password protected to prevent additional software installation. By choosing a strong password, a Hyke administrator can prevent unauthorized software running on the device.

## 4.4 Permutation recording

Since the one time passcode consists of digits displayed to the user, one attack might be to record each possible digit individually, and for an accomplice to reconstruct the passcode audio from the digits for replay. Including letters in the passcode makes this attack more cumbersome, particularly in our user scenario. However, one longer term solution is to append randomly generated dictionary words to the passcode that has to be read for verification. Even if we limit the word selection to a small vocabulary [16], the volume of recorded material needed to launch a permutation attack makes it impractical for our use case.

# 5. IMPLEMENTATION AND EVALUATION

We are currently implementing the Hyke remote attendance system, and have completed some of its critical components. This section will present those implementation details, and our evaluation results. For some components, such as location tagging and speech recognition, we plan to use off the shelf products, and we will present some alternatives.

## 5.1 Biometric verification

Perhaps the most critical component in successfully deploying Hyke is its voice based biometric verification component. Voice based biometric systems are nothing new [5, 9, 17, 20]. However, we were surprised with the absence of an easy to use and open source toolkit that would allow simple integration for system builders. While several companies with voice based biometric systems exist, academic solutions tend to be either too low level for the non-expert [5, 20], or implemented in numerical computing environments that are difficult to integrate with end-to-end systems [9].

Therefore, our goal was to build a complete, end-to-end toolkit that supports an easy interface for speaker verification, and can be used by other projects besides Hyke. The toolkit first trains speaker models using audio recorded from legitimate users, and can later be invoked using a new audio, and an existing model to verify if the utterances came from the same person. To accomplish this goal, we investigated biometric platforms commonly used at the annual NIST speaker recognition competition, including BECARS and ALIZE[3, 5]. We decided to build our toolkit on top of Mistral [11], an open source platform for biometric authentication written in C++. The platform works at the level of feature vectors for various biometric parameters.

The Hyke biometric toolkit provides a Python interface to the underlying platform, and leverages SPro [19], a speech signal processing toolkit, to extract features from audio. Using the Hyke biometric toolkit is simple. Once audio is captured from a user in a supported format (WAV, RAW or the NIST standard SPHERE), it is placed in a designated folder that is processed by the toolkit accordingly. Using training audio supplied by the user, the Hyke biometric toolkit trains models to represent the individual, expressed in terms of variation from a generic 'world' model of all possible audio. As new audio is captured from users, these models are used to verify identity. In speaker verification, this is a 1-to-1 test between an existing model and a newly recorded audio. This audio is scored against the legitimate model, and a confidence score is returned. This score is used to decide whether to accept the user as legitimate or not.

### 5.1.1 *Evaluating speaker verification*

Speaker recognition often involves a large set of configurations that has to be tuned to the operating environment, and we were curious as how well out toolkit would operate on local voices collected from the field. We had an additional constraint that the audio has to be collected over the phone. In telephony, the usable voice frequency band ranges from 300 Hz to 3400 Hz, which is much narrower than the natural frequency of human voice, which ranges from 200 Hz to 7000 Hz. This limits the accuracy of features extracted from the audio signal.

Searching through existing voice data corpuses did not reveal any readily accessible telephone-based datasets collected in developing country contexts. A simple requirement for such a dataset is having several recordings of a single user that can be used to train and test speaker models. As a result, we decided to collect voice data ourselves.

We built an IVR system for collecting voice data, allowing users to call a local number from their phone and record voices. We then posted tasks on Amazon's Mechanical Turk [12], where we limited participation only to workers located in India. We provided workers with the local phone number, and a set of lines to read-out to the IVR. The lines to be read consist of randomly generated digits of various lengths. Since the Hyke system will need to verify identity based on a text independent, limited vocabulary (digit) passcodes, our data collection also focused on this segment.

Our dataset includes voices from 82 unique users, 44 male and 38 female. Each individual has five recordings of length 35 seconds, 20 seconds, 15 seconds, 10 seconds and 5 seconds. In a typical use case, the longest speech sample will be used for training, and the others for testing. The amount of noise in the samples varies, with about 40% of the samples containig noticable noise ranging from a constant background hiss to street noise and songs playing. We believe this sample gives a good representation for office or school based speaker recognition, where recording quality is expected to be average with low to medium noise. We will make this dataset publicly available along with the Hyke biometric toolkit for other researcher to experiment on.

Our evaluations preceded as follows. We first train a world model, which represents a generic speaker model for the system using 25 of our users (13 male and 12 female). Then for each remaining user, we train an individual model using the 35 second sound sample for each. This model represents the unique features of the individual's voice, and will be used for comparison later. We then conducted gender based and combined experiments on various length input. For example, when using a 10 second input of males (which on average corresponds to about 10 to 15 digits read), we compare each 10 second sample of an individual with every individual model available for the male samples. Ideally, a 10 second sample will only match one of those models—the one representing the original user for the sample.

We have represented our results using a Detection Error Tradeoff (DET) curve, which plots the false acceptance rate against the false rejection rate for various choices of threshold scores. The false acceptance rate is the percentage of imposter users that were accepted as legitimate, while the false rejection rate is the percentage of legitimate users that were rejected as impostors. Ideally, the threshold will be set at a point where the false acceptance and false rejection rates are the same, known as the Equal Error Rate (EER). However, Hyke can be configured to minimize one measure, and leverage its fail-over mechanism for decreasing the other.

Figure 1(a) presents a few of the experiments we ran, which have gender specific trials that used various length input samples from users. In all of the runs, the speaker recognition system had an EER of less than 10%. As expected, the longer the input sample, the more accurate speaker recognition is. Figure 1(b) represents our combined experiment of all genders and all sample lengths. This involved over 10,000 comparisons of speaker models. On a desktop with Intel Centrino processor and 2GB of RAM, this task was completed in under 4 minutes. Each one of the samples was compared against every model (of both sexes) available in the sample, and should ideally match only one of them. Each individual had 4 tested samples, and it should ideally generate a total of 4 matches in the experiment. The results of the experiment show the system to have an EER of less than 5%, which is at par with current state of the art according to NIST competitions. In addition, this is sufficient for a host of applications, including the Hyke attendance system.

### 5.1.2 Speech recognition

Speech recognition systems are widely available, both in commercial and open source formats [7]. The Microsoft Windows 7 operating system ships with built in speech recognition software that can be programmed from .NET environments. In addition, Mac OS X also comes with a speech recognition tool. Furthermore, open source projects such as CMU Sphinx [7] provide toolkits for deploying speech recognition. Since our requirement for speech recognition in Hyke is fairly basic, we plan to use an operating system built-in tool.

### 5.1.3 Location and code generation

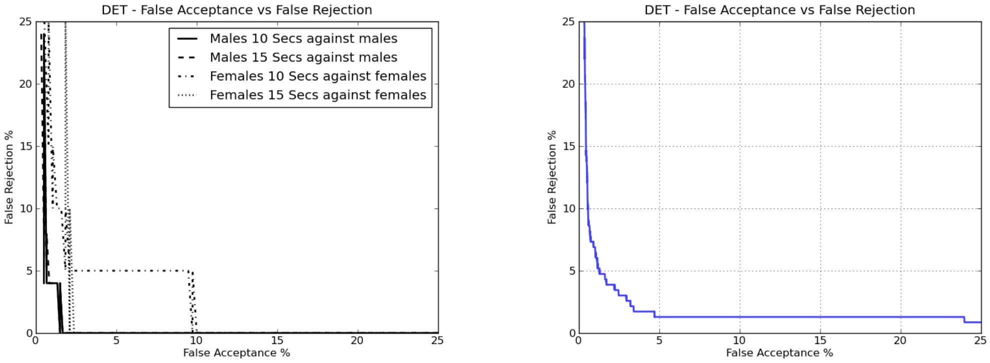
Location tagging in Hyke can be accomplished using a GPS capable phone or by reading a physically secure RFID tag. Once a location ID is determined using either input, the Hyke client displays the one time password. We have implemented a prototype for the server based passcode generation scheme. The server is responsible for maintaining the valid state associated with each location that is currently processing an attendance record.

## 6. RELATED WORK

Biometric identification mechanisms are gaining traction in the developing world. The Unique Identification Number project in India is one well known example that has garnered international attention [21]. The project attempts to create a biometric identification database for all adult citizens of India, with 600 million enrollments expected by 2014. On smaller scales, biometric identification has been used in a number of developing country contexts spanning healthcare, education and business. For example, Weibel et al. looked at using biometric fingerprint identification of mobile pastoralists in Chad [23] while a project in South Africa and Kenya used fingerprints for linking data from geographic health and demographic surveillance systems to data from facility based health management information systems [18].

Voice biometrics systems are becoming increasingly robust, with several commercial offerings. For example, Vodafone Turkey uses a speaker identification systems for enabling self service applications [22], while Bell Canada provides hosted services for speech technologies [4]. The National Institute of Standards and Technology(NIST) has been holding annual speaker recognition competitions for the last 14 years[13]. The state of the art in speaker identification has been steadily improving, with the most recent competition reporting less than 5% error rates for audio recorded over mobile phones. Voice has also been combined with biometric features to identify users [8], and used in local language contexts [17]. Hyke leverages this progress in speaker recognition for building a remote attendance system in challenged network environments. In addition, the Hyke biometric stack incorporates several open source components [11, 14, 19] and provides a simplified interface for non-expert users allowing researchers to incorporate a speaker identification technology in their projects.

Biometric attendance tracking has also been considered in several projects. A recent project looked at using fingerprints for managing attendance in Indian health care pro-



(a) Similar gender tests

(b) Combined gender test

**Figure 1: DET's for Hyke's voice biometrics stack**

grams [15]. The system uses off the shelf components including a netbook computer, finger print reader and a mobile phone for sending SMS messages to a central server. Earlier work in Malawi also uses fingerprints for identifying AIDS patients on medication [24]. As described in section 2, the Seva Medir school system is another example of a semi-automated biometric attendance system that uses pictures for tracking presence. There are three main advantages Hyke offers over these and similar biometric attendance systems. First, it reduces the end user component required to track attendance, allowing organizations to deploy the system in remote areas where its not feasible to distribute computer-based systems. Second, Hyke is designed with independent location verification as a first order concern. This is especially important in environments where its not feasible to have trusted administrative staff at all times. Finally, voice offers a less invasive form of biometric than fingerprints or pictures. In addition, the high penetration of cellular coverage in developing regions allows for a wide deployment of voice based technologies.

## 7. CONCLUSION AND FUTURE WORK

Hyke was designed to provide a simple and cost effective attendance tracking mechanism for remote locations. Our evaluation of its biometric stack suggests that it could be used efficiently in many of these environments with an appropriate fall-back mechanism. In the case of the Seva Mendri school system, an automated and real-time attendance tracking system will significantly reduce the burden on human operators, and errors that get introduced. The feedback we get from the non-profit organization as well as the research community will be important in refining the design for practical use.

## 8. REFERENCES

[1] Seva Mandir. www.sevamandir.org/.

[2] Biometric in 40,000 schools. eSSL Biometric Secuwatch, 2007. http://secuwatch.wordpress.com/biometric-in-40000-schools/.

[3] BECARS Library and Tools. `perso.telecom-paristech.fr/chollet/becars/index.php`.

[4] Bell Canada Voice ID, 2011. `www.bell.ca/home`.

[5] J.-F. Bonastre, F. Wils, and S. Meignier. ALIZE, a free toolkit for speaker recognition. In *ICASSP'05, IEEE*, 2005.

[6] N. Chaudhury, J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers. Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, 20(1):91–116, 2006.

[7] CMUSphinx Speech Recognition. `cmusphinx.sourceforge.net/`.

[8] A. Das, O. Manyam, M. Tapaswi, and V. Taranalli. Multilingual spoken-password based user authentication in emerging economies using cellular phone networks. In *Spoken Language Technology Workshop, 2008.*

[9] H. Do, I. Tashev, and A. Acero. A new speaker identification algorithm for gaming scenarios, 2007.

[10] E. Duflo, R. Hanna, and S. Ryan. Monitoring works: Getting teachers to come to school. *CEPR Discussion Papers*, 6682, 2008.

[11] Mistral Plateforme open source dï£¡authentification biometrique. `mistral.univ-avignon.fr/en/`.

[12] Amazon Mechanical Turk. `www.mturk.com`.

[13] NIST Speaker Recognition Evaluation . `www.itl.nist.gov/iad/ mig/tests/sre/`.

[14] SPHERE Speech File Manipulation. `www.itl.nist.gov/iad/mig/tools/`.

[15] M. Paik, N. Samdaria, A. Gupta, J. Weber, N. Bhatnagar, S. Batra, M. Bhardwaj, and W. Thies. A biometric attendance terminal and its application to health programs in india. In *NSDR 2010*, pages 4:1–4:6. ACM.

[16] F. Qiao, J. Sherwani, and R. Rosenfeld. Small-vocabulary speech recognition for resource-scarce languages. In *ACM DEV 2010.*

[17] R. Rajeswara, R. A. Nagesh, K. Prasad, and K. E. Babu. Text-dependent speaker recognition system for indian languages.

[18] A. Serwaa-Bonsu, A. J. Herbst, G. Reniers, W. Ijaa, B. Clark, C. Kabudula, and O. Sankoh. First experiences in the implementation of biometric technology to link data from health and demographic surveillance systems with health facility data. *Global Health Action*, 3, 2010.

[19] SPro Speech Signal Processing. `www.irisa.fr/metiss/guig/spro/`.

[20] A. Subramanya, Z. Zhang, A. C. Surendran, P. Nguyen, M. Narasimhan, and A. Acero. A generative-discriminative framework using ensemble methods for text-dependent speaker verification, 2007.

[21] Unique Idengification Authority of India, 2011. `www.uidai.gov.in/`.

[22] Vodafone Turkey, 2011. `www.vodafone.com.tr/`.

[23] D. Weibel, E. Schelling, B. Bonfoh, J. Utzinger, J. Hattendorf, M. Abdoulaye, T. Madjiade, and J. Zinsstag. Demographic and health surveillance of mobile pastoralists in chad. *Geospat Health*, 3(1):113–24, 2008.

[24] K.-L. Yu, C.-C. Chen, W.-S. Chang, H. Juma, and C. S. Chang. Fingerprint identification of AIDS patients on ART. *The Lancet*, 365, 2005.