Odyssey 2012
The Speaker and Language Recognition Workshop
25-28 June 2012, Singapore

The RATS Radio Traffic Collection System

Kevin Walker, Stephanie Strassel

Linguistic Data Consortium
University of Pennsylvania, Philadelphia, PA USA
{walker|strassel}@ldc.upenn.edu

Abstract

The DARPA RATS Program focuses on the development of new technologies for identifying and processing speaker-to-speaker communications over degraded radio channels. In order to build a corpus to address this research question, we developed a system that takes a clean source signal and transmits it over eight different radio channels, where the variation from channel to channel results in a range of degradation modes. Each channel included in the collection system has unique characteristics targeting different modulation types, different carrier channel bandwidths, and different operating bands.

1. Introduction

The goal of the DARPA Robust Automatic Translation of Speech (RATS) Program is to develop technologies capable of performing speech activity detection (SAD), language identification (LID), speaker identification (SID) and keyword spotting (KWS) in potentially speech-containing signals received over communication channels that are extremely noisy and/or highly distorted. The first phase of the program evaluates these technologies in five languages – Levantine Arabic, Farsi, Urdu, Pashto and Dari – on open source data that nonetheless represents real-world operational scenarios. Later phases of RATS incorporate secondary evaluations using classified "real world" data.

As with other large human language technology evaluation programs, RATS requires a multi-pronged, integrated and creative data collection approach to satisfy the combined demands of relevant data, language diversity, rich annotations and high volume. The unique challenge for RATS corpus development is the additional need to produce training, development and evaluation data with extremely challenging acoustic properties while simultaneously maintaining efficient, cost-effective and scalable collection and annotation methods. To address this challenge we have developed a "collect clean, broadcast dirty" methodology. In this model we employ well-tested collection methods [1, 2] to produce clean-signal source data for each language and task. Manual annotation (speaker and language auditing, segmentation and time stamping, transcription) is performed on this non-degraded signal to assure efficiency and accuracy. We then introduce the desired noise properties into live recordings and content rebroadcasts by means of a Multi Radio-Link Channel Collection System. The system includes

- A set of target signal transmitters/transceivers
- A set of interference signal transmitters,
- A set of listening station receivers,
- Signal collection and digitization apparatus

In this paper we discuss design and implementation of the Multi Radio-Link Channel System, with special attention paid to the motivation behind the radio channels included in the system, implementation challenges and solutions, and the resulting data.

2. System Design

Many modern two-way communications systems used in urban environments by law enforcement and emergency services have transitioned to digital modes like APCOP25, MOTOTRBO, or TETRA. These established systems rely on extensive repeater installations to maximize service area, utilize trunking or packet switching techniques in order to maximize the number of concurrent users, and in some cases employ encryption in order to control information exposure. It is important to note that these systems are thoroughly engineered, may be cross-connected with the public telephone network and computer networks, and represent an established "installation". They aren't designed for ad hoc communications, and their operational parameters are clearly defined – their designers hope that the systems deliver predictable level of service. As such, internally these networks may be of relatively less interest when looking at questions of system performance in the face of unpredictable variation.

As a contrastive example, consider the Air Traffic Control system used currently in the US and around the world. This system is also well connected, has considerable redundancy, and must provide at least some level of quality of service. Where ATC differs from the communications radio installations used by law enforcement (for example) is in the fact that its users are not predictable in the same way. All pilots, whether operating large commercial airplanes, corporate jets, or private airplanes, depend on instructions, assistance, and direction from air traffic controllers. It may be the case that large airports provide separate channels depending on vehicle type, but even within a given category, we see variations in radio equipment performance. ATC traffic is typically not encrypted, uses amplitude modulation to avoid FM capture effect, relies heavily on the VHF band, and has a constantly changing set of users. The system has to be open and flexible in order to accomplish its mission – it will still need flexibility and robustness to variation across users even as it modernizes.

When developing the data collection system for RATS, we envisioned a set of radio communications scenarios with significantly greater variation and degradation than one would find in the highly controlled communications networks, and even greater than what one would find in the Air Traffic Control model. We imagined a scenario in which interlocutors were located in an indeterminate terrain, utilizing equipment of unknown origin and unknown operation parameters, and whose communication was being monitored by a third party whose situation vis-à-vis the interlocutors was also not know a priori. In this scenario, the communications path between interlocutors might be of varying quality, and, furthermore, the communications path between the interlocutors and the monitor would almost certainly be of varying quality, and would be highly dependent on environmental, atmospheric, and structural factors. Rather than taking communications among a police force or a control tower and a Cessna as our target, we were more interested in looking to something like the communications among a group of amateur radio operators, or the handheld radio communications of a team playing Capture-The-Flag in a local state forest. Consequently, we targeted a channel model of a single Shared Channel used in half-duplex mode. This model implies that operators may either transmit or receive, but may not do both simultaneously; furthermore, attempts of more that one operator to utilize the channel at the same time result in partial or total communication failure.

One type of variation encountered in radio communications is in speech intelligibility. Our collection system is interesting because it produces samples whose intelligibility varies both in terms of degree and in terms of type. The effects of signal degradation present differently on a single sideband HF radio channel than they do on an FM UHF channel, and both fail in different ways than a spread spectrum system operating in the SHF band. In other words, while a phenomenon like multipath fading can be seen across all bands, its effect - that is to say, the effect on the contents of the channel - varies greatly depending on the operational parameters of the radio equipment in question.

A critical first step in developing our collection platform was a review of extensive work that went into the development of the Tactical Speaker Identification Speech Corpus (TSID) [3]. This corpus, produced at MIT Lincoln Labs by Doug Reynolds and Gerald C. O'Leary was used as a model for our collection protocol. In particular, the selection and configuration of radio transceivers, the use of Map Task scenarios to elicit transactional, task oriented speech, and the selection and organization of session information were all extremely helpful guidelines. The TSID Corpus was collected at Fort Bragg; it employed a central fixed transmitter location and a distribution of receiver sites at a range of distances from the transmitter site. One particular point of interest that we would like to incorporate into future collection is the fact that the topography of the collection location had significant variation.

A second prior work that helped to inform our development of the RATS collection system was the SPINE2 data collection effort [4]. Arcon Corporation collected the data at Fort Irwin and Fort Knox in 1999 for the Digital Voice Processing Consortium. SPINE2 also emphasized task driven communication across radio channels in ways that were unpredictable, highly variable, and representative of real world operational activity.

Finally, our system development efforts were informed by a series of initial outdoor Signal Quality surveys performed by LDC staff. These surveys provided real world experience of the significant impact of terrain, buildings, and proximity to metal signs & structures, and elevation on signal quality and radio performance. These initial surveys informed our decision to focus on carrier wavelength, carrier bandwidth, and modulation as variables to be represented in the collection.

3. System Implementation

3.1. Overview

The collection system includes a set of *target signal transmitters/transceivers*, a set of *listening station receivers*, and *signal collection and digitization apparatus*. It is used to process both live sessions and rebroadcast sessions. Recordings selected for rebroadcast consist of dialogues and conversational content. In order to simulate Push-To-Talk, the transceivers are keyed under computer control. The control computer is connected to the transceivers via two routes: the audio signal route and the control route. The audio signal route originates with a *Lynx Studio Technologies AES16e* digital audio interface connected to a *Lucid 88192* Digital to Analog converter. The DAC unit is connected to matrix mixer, which provides analog audio to the input of each transceiver.

The matrix mixer audio output is 600-ohm line level, and must be impedance matched to the audio input for each individual transceiver. Additionally, independent transformers are used to electrically isolate the transceivers from the control computer and audio processing hardware. In order to control the push to talk functions of the transceivers, the control computer is connected to an SPDT relay bank via TCP/IP. The relay bank includes eight relays, each of which is used for one transceiver (each with a customized wiring harness). The control computer monitors the signal strength of the input source recording and keys the transceivers based on a custom signal activity detector.

3.2. Estimation Of System Induced Signal Degradation

In order to estimate the signal-to-noise ratio of each retransmission, we performed a modified *signal* to *noise+distortion* measurement for each channel. We transmitted a 400ms, 1KHz sine test tone through all eight channels and recorded the output. The test tone was generated digitally with an RMS power measure of -5dBFS. The dBu level at each transceiver was adjusted to provide optimal analog levels for the respective input circuit; this gain adjustment ranged from 0 to -12dB depending on the handset. This estimate should not be considered a true SINAD measurement. It illustrates the degree of signal modification introduced by each channel. These measurements, summarized in Table 1, serve as an initial step towards a more robust assessment of retransmission channel characteristics; they demonstrate the ratio of the intended signal to non-signal components found in each channel recording. The procedure used to generate these values was as follows:

I. Generate 1KHz test input signal.
II. Transmit the test signal across the radio channels and record the output from the receivers.
III. Measure the rms power of resulting recordings. This measurement represents the signal transferred from transmitter to receiver plus any introduced noise and distortion (SND).
IV. Apply a narrowband (high Q-factor) 1khz notch filter to each retransmission channel recording. Measure the rms of each resulting sample. This measurement represents noise and distortion (ND).
V. Subtract ND from SND to produce an estimate of the non-signal components present on each radio channel.

Retrans Channel	SND (dBFS)	ND (dBFS)	SND - ND (dBFS)
1	-9.03	-26.05	17.02
2	-9.03	-28.37	19.34
3	-8.87	-23.64	14.77
4	-11.06	-22.61	10.85
5	-10.11	-27.33	17.22
6	-9.51	-24.72	15.21
7	-9.58	-16.48	6.9
8	-17.22	-22.06	4.87

Table 1: Retransmission Introduced Noise and Distortion

The types of signal degradation found in the retransmission recordings include pitch shifts, ring modulation, long time scale amplitude variation, non-linear, intermittent fading and harmonic attenuation.

3.2 Transmitters and Receivers

The Target Signal Transmitters and Transceivers cover a variety of radio channel configurations. We included AM, Narrow FM, Wide FM, Single Side Band, and Frequency Hopping Spread Spectrum handsets. Two of the transmitters were operated in the HF band, one in the VHF band, three in the UHF band, and two in the SHF band. The AM, FM, and SSB handsets employed narrow (<5Khz) channel bandwidths. We chose to emphasize relatively low radiated power handsets for two reasons – firstly, to be consistent and representative of what would typically be found in the field, and secondly to accommodate regulatory restrictions. The effective radiated power of the transceivers included in our system range from 0.5W to 12W, depending on modulation and operating band. The transceivers were all configured to generate RF field strength ranging from 750 to 10000 µV/meter @ 3 meters.

All of the transceivers were equipped with low gain, omnidirectional, vertically polarized, monopole aerials, each with an electrical length of ¼ wavelength of the transmit frequency for a given transceiver. The characteristics of each transceiver are summarized in Table 2.

Transceiver	Band	Frequency	Modulation	Channel Separation	Max Deviation	ERP
Motorola HT1250	UHF	456MHz	NFM	12.5kHz	2.5KHZ	4W
Midland GXT1050	UHF	462MHz	NFM	6.25kHz	2.5kHz	0.5W
Icom IF-F21GM	UHF	462MHz	NFM	6.25kH	2.5kHz	1W
Galaxy DX2547	HF	27KHz	SSB	10KHz	NA	12W PEP
Icom IC-F70D	VHF	151MHz	NFM	11.25KHz	2.5KHz	2W
Trisquare TSX300	SHF	900MHz	FHSS	NA	NA	1W
Vostek LX3000	SHF	2.3GHz	WFM	200KHz	25KHz	2W
Magnum 1012 HT	HF	27MHz	NFM	10KHz	2.5Khz	1W

Table 2: Transmitter Bank

We built a listening post to capture signal from the target transmitters. The listening post included a bank of wideband receivers, each of which was tuned to correspond with target transceivers. A central computer, either via RS-232, TCP/IP, or contact closure, depending on model, directly controlled all of the receivers. All of the receivers were equipped with omnidirectional, wideband, vertically polarized, monopole aerials. The selection of omnidirectional antennas was intentional; we wanted to stress each receiver's ability to reject reflected, off axis signals. Our goal was to maximize the potential for multipath degradation of the source transmission, and to avoid attenuation of non-target transmissions. As with the transceiver bank, the antennas were electrically matched with the intended carrier frequency in most cases. In the case of the Trisquare FHSS transceiver, we used a paired transceiver as its partner in the receiver bank. We considered the possibility of configuring a high speed scanner to sweep all of the possible frequencies used by th7e Trisquare transceiver, but decided that since these particular handsets were designed to operate in tight coordination with each other, that the correct solution for the purposes of this data collection was to rely on the second member of the transceiver pair. The characteristics of each receiver are summarized in Table 3.

Receiver	Band	Frequency	Modulation	IF Bandwidth
AOR AR5001D	UHF	456MHz	NFM	3KHz
AOR AR5001D	UHF	462MHz	NFM	6KHz
TenTec RX400	UHF	462MHz	NFM	15KHz
Icom IC-R8500	VHF	151MHz	NFM	6KHz
Icom IC-R75	HF	27MHz	SSB	9KHz
Trisquare TSX300	SHF	900MHz	FHSS	NA
Vostek VRX24	SHF	2.3GHz	WFM	200KHz

Table 3: Receiver Bank

These receivers were selected because of their reliability, extreme flexibility, and well-developed support for computer integration. The TenTec and AOR receivers in particular have proven to be excellent choices for a lab environment because of their facility for managing configuration settings in a transparent and highly detailed manner.

The mapping of transmitter to receiver (summarized in Table 4) was organized to make comparisons of within groups of similar receivers possible. In some cases, the choice of transmitter/receiver pair was predetermined by the transmitter technology itself (e.g. the FHSS handsets). On the other hand, in the case of the VHF and UHF transceivers, we had more flexibility in making the assignment. In these cases, our focus was on varying the receiver configuration as well as the transceiver power output level. In some cases, we configured the receiver so that it was operating on the margin of its failure mode. For example, in the case of the TenTec RX400, operating on the UHF band with a wide intermediate frequency state, the tuned frequency was offset from the transceiver to a sufficient degree that the FM demodulator stage selectivity and capability to stay locked on the transmit frequency was at its extreme margin.

In addition, an external process was used to gradually perform random shifts the target receive frequency within a narrow window. The tonal distortions found in audio from this channel are caused by the receiver FM detector continuously attempting to lock onto the transmit frequency.

RF Channel	Transmit - Receive Pair		Configuration Details
UHF NFM	Motorola HT1250	AOR AR5001D	Receiver: Dual Frequency Mode, 50KHz offset
UHF NFM	Midland GXT1050	AOR AR5001D	Receiver: noise reduction enabled
UHF NFM	Icom IC-F21GM	TenTec RX400	Receiver: 3KHz offset relative to transmitter
VHF NFM	Icom IC-F70D	Icom IC-R8500	Transceiver: Companding function enabled
HF SSB	Galaxy DX2457	Icom IC-R75	Tonal variation caused by SSB drift
SHF FHSS	Trisquare TSX300	Trisquare TSX300	
SHF WFM	Vostek LX-3000	Vostek VRX24	
HF NFM	Magnum 1012 HT		

Table 4: Transmitter/Receiver Pairings

3.3. Signal Collection And Digitization Apparatus

The listening post recording computer is a standard Linux workstation with a low latency kernel, an eight channel *Digigram VX882e* digital audio capture card, and dedicated capture hard disk drives for each audio channel. The host operating system is Ubuntu 10.04, and we made use of the ALSA audio framework and API. The recordings were made using Ecasound. Due to the fact that the receivers presented a range of audio outputs, each receiver output was impedance normalized to 600-Ohm line level. The audio recordings were captured as 48KHz, 16bit, linear PCM RIFF wav files; they were then downsampled to 16khz and converted to flac. The recording computer also performed double duty as the receiver control computer. In order to create a control path between the recording computer and the receiver bank, we used a *Comtrol* multiport RS-232/TCP-IP bridge. Each receiver with a serial port was connected to the multiport bridge and assigned a specific IP port. The multiport bridge was configured to have a direct Ethernet connection to the recording computer, and the recording computer was able to access and communicate with a given receiver by connecting to its assigned port. Custom Perl modules were written implementing the AOR, Icom, and Ten Tec control protocols; these modules were used to configure the receiver's frequency, mode, and operating functions such as automatic gain control, IF bandwidth, noise reduction level, and pass band via RS-232 or TCP/IP. The recording computer also includes a MySQL database that contains receiver configuration information, recording session information, and job control status information.

Because we rely on two separate computers in separate locations to handle the retransmission process, the resulting recordings are offset by some amount of time relative to the source recordings. The computers are synchronized at a coarse level due to the fact that they both rely on a single ntp server; however, the fine level required for synchronization of digital audio is not currently handled in real time. Our approach is to ensure that the retransmission recordings bracket the source recording playback. Once retransmission has completed we use cross correlation to establish the alignment point between the source file and the retransmission recordings. The alignment tool that we use is called *skewview* (http://labrosa.ee.columbia.edu/projects/skewview), developed by Dan Ellis of Columbia University. The tool is implemented in Matlab; it generates a running stream of correlation coefficients. Each cross correlation maxima represents a single alignment hypothesis for a short section of the reference and test recordings. By calculating a running set of alignment hypotheses, *skewview* allows us to check for clock drift between the reference and test recordings. The final decision to use the *Lynx Studio Technology AES16e* audio interface for playback and to use the *Digigram VX882e* audio interface for recording were validated by checking sample retransmission sessions for clock drift using *skewview* (a different audio interface had to be excluded from the system specifically because it was unable to produce long duration playback without introducing an unacceptable level of clock drift). The current system produces retransmission recordings with a consistent offset relative to the source audio, without introducing time skew between them. Each retransmission session is processed with *skewview* as part of the normal RATS data preparation pipeline.

We have considered the possibility of locking the transmit control computer to the recording computer via wordclock; however, the distance between the two stations makes this impractical. Furthermore, one of the underlying assumptions of the motivation for this system is that the subsystems might be in motion; tethering the transmit and receive systems to one another or to a master clock would limit their mobility.

4. Challenges & Solutions

Three challenges that confronted us during deployment and operation of the retransmission system were the supply of adequate, clean power to the radio equipment; creation of a mechanism for synchronizing push-to-talk activity across the set of transceivers; and detecting failures in the transmission process.

When the system was initially deployed, we used a combination of independent "wall wart" DC power transformers and handset power cradles to supply power to the transceivers. After operating the system in an experimental mode, it became clear that this configuration was impractical. It was difficult to manage such a large number of power supplies, it caused problems for the rechargeable batteries supplied with certain transceivers, and in some cases an unacceptable level of ground loop hum was injected into the signal path due to the poor quality of some of the manufacturer provided power supplies.

In order to improve the system's operation with respect to transceiver power, we tried two approaches. The first step was to replace all of the rechargeable battery packs with off-the-shelf battery eliminators. Since the retransmission system was required to operate for extended periods of time, the manufacturer provided battery packs were overtaxed and did not last very long – in some cases, the duty cycle demanded by the retransmission schedule kept the battery packs from charging completely, resulting in marginal performance. Our attempt at using an off-the-shelf battery eliminator was a disappointment; we purchased a battery eliminator for the *Motorola HT1250* handset, but poor build quality and output voltage deviation rendered it unusable.

The second approach proved to be more successful. We chose was to equip each transceiver with a high quality, dedicated, variable output voltage regulator. We eliminated all individual external AC/DC plug-in transformers and powered all of the transceivers using a single Alinco 35amp 12VDC benchtop power supply. The common power supply was connected to a customized power distribution panel, which provided power for all of the transceivers. Because the input voltages and power requirements vary from transceiver to transceiver, the connection between the power distribution panel and the transceivers is mediated by a set of dedicated configurable output voltage regulators which provide power with the correct voltage and current requirements for each transceiver. The variable output DC voltage regulators that we are using are sold by Dimension Engineering (http://www.dimensionengineering.com/anyvolt3.htm); the AnyVolt3 Voltage regulator is capable of stepping a 12VDC input to any output from 3 to 24 VDC at up to 3 amps current draw. This exceeds the requirements of the handsets that we are using for our system. Consolidation and refactoring of the power distribution requirements of the collection system has made the system easier to maintain and has improved the longevity, consistency, and overall performance of the radio equipment.

The challenge of synchronizing the push-to-talk functionality of the transmit systems transceivers was addressed by installing an RDL Signal Level Controlled Relay along with a National Control Devices computer controlled Multi-SPDT Relay Bank into the system. The SPDT Relay bank (http://www.iorelay.com) provides eight independently controllable relays and eight independent ADC channels which provide input voltage potential level relative to a +5VDC reference, with 8-bits of precision. The Relay Bank employs a simple, well-documented control protocol, depicted in Figure 1. We developed a Perl module which implements the protocol and handles all socket transactions; the session management software which runs on the transmit station control computer spawn a separate subprocess which controls the SPDT Relay component by monitoring the signal strength of the source recording being retransmitted. The RDL Signal Level Relay combines an audio input with variable sensitivity with an SPDT relay. When the audio level input crosses a preset threshold, the relay closes, providing a -5VDC potential. This transition pulls down the level of one of the ADC channels on the Relay Bank. When the audio level drops below the threshold, the relay transitions back to 0VDC potential. The relay is configured to use a fast attack, long sustain, and gradual release. By utilizing a fast attack and slow release, the system provides generous padding around all detected utterances. The signal onset threshold is -30dBu, and the relay uses a 25ms attack time, a minimum 5 second sustain, and a minimum 2second release duration. Finally, the signal monitor pre-fetches the audio signal so that it is always at least 1 second advanced relative to the audio stream being fed to the transceiver bank.

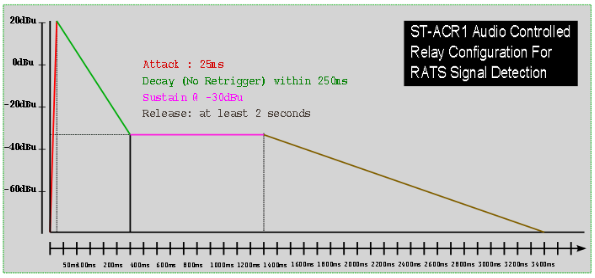

Figure 1: RDL Signal Level Relay Response Graph

A final challenge was the detection of transmission failures that can result in audio that contains only silence or only noise, rather than the targeted noisy speech. A variety of approaches were considered, but it was important to develop a method that was both efficient and scalable given the large volume and compressed timeline for data deliveries. Ultimately we settled on a fairly simple solution that performs very well for most of the transmission channels. First, we calculate the root mean square (RMS) moving average for each channel after retransmission. We also compare frame-based RMS envelopes among the original clean source recording and the corresponding retransmission recordings. We then look for and enumerate cases where either there is no signal at all, or the entire recording had little or no variance in frame-to-frame RMS power (indicating that the carrier signal on the given transmitter channel had never engaged over the full duration of the recording session). Such recordings are excluded from the downstream pipeline.

5. Resulting Data

The RATS Radio Channel system has been used to process thousands of hours of audio in all five of the RATS target languages – Levantine Arabic, Farsi, Dari, Pashto and Urdu – plus multiple imposter languages for Language ID, as well as additional English data for supplemental Speech Activity Detection training. This audio has been drawn from a variety of sources, including existing LDC corpora as well as new data collection.

Dataset	Language	Hours of Source Audio
RATS New Telephone Speech Collection	Pashto	393
	Urdu	187
	Farsi	75
	Dari	2.7
	Levantine	245

Table 5: New RATS Telephone Collection at LDC

To support the requirement for narrowband speech from many unique individuals, LDC employs speakers of each RATS language to recruit other native speakers from within their social networks who consent to having their speech recorded for inclusion in the corpus. Each participant makes one or more call, and the calls are recorded via LDC's existing Conversational Telephone Speech collection system. Some recordings consist of general conversation, while other sessions involve a series of scenario-driven language games like *Twenty Questions*, *Battleship*, *Scavenger Hunt* which are used to encourage the use of transactional speech and short-duration interchanges. Table 5 summarizes the data collected to date under this protocol.

To supplement new data collection we have also drawn heavily from several existing LDC corpora. First, to provide additional LID data in both the RATS target languages and multiple imposter languages, we selected segments from previously exposed NIST LRE Evaluation data sets [5]. We also received assistance from the Brno University of Technology, Faculty of Information Technology Speech Group in identifying additional narrowband recordings from the LRE09 VOA data set [6]. Data drawn from the various NIST LRE data sets is summarized in Table 6.

Dataset	Language	Hours of Source Audio
NIST LRE Test Sets (Various)	Farsi	7.2
	Urdu	5.6
	Dari	3.7
	Pashto	3.6
	imposter languages combined	276
VOA3 Narrowband	Urdu	12
	Dari	12
	Pashto	11
	Farsi	9
LRE09	Pashto	1.7
	Dari	1.7

Table 6: NIST LRE Data Used in RATS LID

Finally, three existing telephone speech corpora, originally collected by LDC to support automatic speech recognition and language identification systems, were selected for retransmission in order to gain additional coverage of English, Levantine Arabic, and Farsi (Table 7).

Dataset	Hours of Source Audio
Fisher Levantine	275
Fisher English	168
Callfriend Farsi	85

Table 7: Existing Telephone Speech Corpora

Table 8 summarizes the full set of data, regardless of data source provided by LDC to support the LID evaluation in Phase 1 of RATS.

Partition	Language	Files	Hours
TEST	Levantine	878	29.3
	Farsi	1009	35.1
	Dari	237	8.5
	Pashto	866	29.5
	Urdu	887	31
	imposter languages combined	2470	161
TRAIN	Levantine	3849	128.3
	Farsi	399	14.6
	Dari	133	4.9
	Pashto	2574	86.2
	Urdu	1717	58.3
	imposter languages combined	2690	141.8

Table 8: LID Data

Beyond the data described above, additional data collection is required to support the SID evaluation since ground truth speaker identity is not known for several of the existing corpora used in other RATS evaluation tasks. LDC and its partner sites are currently engaged in collection of additional telephone conversations from up to 150 speakers in each of the five RATS languages, targeting 10 conversations per speaker. Speaker auditing is ongoing, with the first SID evaluation scheduled for June 2012. Data collected to date for SID is summarized in Table 9 below, along with the expected number of unique speakers.

Language	Expected Unique Speakers	Files	Hours
Levantine	146	1099	223.1
Farsi	39	20	4
Dari	14	20	4.1
Pashto	95	1173	237.2
Urdu	106	511	103.5

Table 9: SID Data

The linguistic resources described in this paper have been distributed to RATS performers as training, development and evaluation data. We will wherever possible distribute the data more broadly, for example to Linguistic Data Consortium Members and Licensees via publication in the LDC catalog. Upon sponsor request some subsets of data may be reserved for use within RATS only. The first set of RATS data is targeted for release in LDC's catalog in early 2013.

6. References

[1] Christopher Cieri, Joseph P. Campbell, Hirotaka Nakasone, David Miller, Kevin Walker. 2004. *The Mixer*

Corpus of Multilingual, Multichannel Speaker Recognition Data. Proceedings of LREC 2004: Fourth International Conference on Language Resources and Evaluation, Lisbon.

[2] Christopher Cieri, David Miller, Kevin Walker. 2002. *Research Methodologies, Observations and Outcomes in (Conversational) Speech Data Collection*. HLT 2002 The Human Language Technologies Conference, San Diego, CA, March 2002.

[3] David Graff, Douglas Reynolds and Gerald C. O' Leary 1999. *Tactical Speaker Identification Speech Corpus (TSID)*. Linguistic Data Consortium, Philadelphia.

[4] Astrid Schmidt-Nielsen, et al. 2001. *Speech in Noisy Environments (SPINE2) Part 2 Audio.* Linguistic Data Consortium, Philadelphia.

[5] NIST. 2011. Language Recognition Evaluation Website (*www.itl.nist.gov/iad/mig/tests/lre/*.)

[6] Christopher Cieri, Linda Brandschain, Abby Neely, David Graff, Kevin Walker, Chris Caruso, Alvin F. Martin, Craig S. Greenberg. 2009. *The broadcast narrow band speech corpus: a new resource type for large scale language recognition*, In INTERSPEECH-2009, 2867-2870.