Employing Phonetic Speech Recognition for Language and Dialect Specific Search

Corey Miller
UMD CASL
7005 52nd Avenue
College Park, MD
20740
cmiller6@umd.edu

Rachel Strong
UMD CASL
7005 52nd Avenue
College Park, MD
20740
rstrong1@umd.edu

Evan Jones
UMD CASL
7005 52nd Avenue
College Park, MD
20740
jone1072@umd.edu

Mark Vinson
UMD CASL
7005 52nd Avenue
College Park, MD
20740
mvinson@umd.edu

Abstract

We discuss the notion of language and dialect-specific search in the context of audio indexing. A system is described where users can find dialect or language-specific pronunciations of Afghan placenames in Dari and Pashto. We explore the efficacy of a phonetic speech recognition system employed in this task.

1   Introduction

The Audio Gazetteer hotspotting tool was developed by MITRE (2012) and employs the Nexidia phonetic speech recognition engine (Gavalda and Schlueter, 2010) in several languages, including Dari (the Afghan variety of Persian) and Pashto, the two main languages of Afghanistan. These languages are both members of the Iranian language family and share a number of phonetic characteristics (Miller et al., 2013). This tool enables a user to load audio clips and to search them for words contained within them using one of three methods: the Dari or Pashto alphabets, a Romanization scheme, or phonetics in SAMPA (Wells, 1997). Such a search will yield each starting timepoint in an audio file where the system has identified the term being searched, along with a number between 0 and 100 indicating the level of confidence the system has in its determination. While terms of any kind can be searched, the system provides additional mapping capabilities for placenames.

Audio hotspotting, also known as keyword spotting or audio indexing, is a form of information retrieval employing speech recognition that is used for quickly identifying passages of interest within audio files. It can be used to identify calls of interest in call centers, or to explore reports of natural disasters or political crises in the media. There are two main approaches to audio hotspotting; one involves speech-to-text (STT), also known as large vocabulary continuous speech recognition (LVCSR), and the other employs phonetic speech recognition.

STT ingests speech and outputs orthographic text. To do this, it requires language-specific acoustic and language models mediated by a pronunciation model or dictionary that maps words to phonetic forms. The output text transcript can then be mined for terms of interest. Raytheon's BBN Broadcast Monitoring System is an example of such a system (Raytheon, 2012). One liability of this approach is the need to establish the vocabulary, upon which the language and pronunciation models depend, upfront. That means that one cannot easily search for terms that have not been programmed into the system beforehand. This is an especially challenging impediment when confronting natural disasters and political crises in regions with towns and personalities whose names are "out of vocabulary" (OOV).

Phonetic speech recognition uses language-specific acoustic models directly; allowing users to query phonetic strings, possibly with the aid of a pronunciation model allowing orthographic search. The ability to query phonetic strings removes the OOV problem; any string that can be composed of the phonemes of a particular language can be searched. While this technology is useful for keyword spotting, it cannot be used to generate a meaningful orthographic transcript of speech, due to its lack of a language model.

Our purpose is to explore the feasibility of using phonetic speech recognition technology to explore subtle dialect and language differences, with the ultimate aim of enabling language or dialect specific search. In such a scenario, a user is not simply interested in finding a particular term of interest, he is also interested in the sociolinguistic characteristics of the speaker of that term of interest.

Various researchers have performed promising experiments using STT to explore phonetic variation. These experiments utilize STT in forced alignment mode; that is, given a pre-existing orthographic transcript, they ask the recognizer to focus on deciding which pronunciation among a finite set supplied by the researcher maps best onto particular audio exemplars. Fox (2006) used this technique to examine several realizations of syllable-final /s/[1] in Spanish including [s], [h] and deletion, while Wester et al. (2001) explored variable deletion of /n/, /r/ and /t/ in Dutch, as well as schwa-insertion and deletion. Both demonstrated promising agreement between the STT-based approaches and human coding.

In contrast, the phonetic speech recognizer employed here requires neither an orthographic transcript, nor a predetermined set of phonetic variants from which to choose. For that reason, we felt it offered a flexible platform from which to explore phonetic variation, and thus enabled employing knowledge of that variation to perform dialect and language-specific search for Dari and Pashto.

2   Data collection and transcription

We developed an interview protocol consisting of three components: a sociolinguistic background interview, a map task and a word list. This interview was designed to elicit Afghan placename data from Afghans residing in the United States whose native language was either Dari or Pashto. Speakers bilingual in Dari and Pashto were interviewed in both languages sequentially. Seven Dari and three Pashto interviews, comprising approximately six hours in total, were digitally recorded and later downsampled to 16 MHz with 16-bit precision.

The purpose of the sociolinguistic background interview was to establish the language and dialect profile of each speaker. Where possible, it was conducted in the speaker's native language, and established the location and duration of each place where he or she resided. In addition, the interview established the location and language of instruction of each school attended, as well as the language and dialect used with family members and friends. The interview inquired about all the languages and dialects both spoken and understood by the speakers.

The purpose of the map task was to gather subjects' pronunciations of placenames in Afghanistan in a casual style. A large colored map of Afghanistan, using native lettering, was placed before the subjects and they were asked to explain in Dari or Pashto how to get to and from various points.

The final part of the interview involved reading a word list in Dari or Pashto containing the names of over 200 placenames, including provinces, provincial capitals, other large towns, administrative divisions, regions, mountain ranges, passes, bodies of water, airports and deserts. In Pashto interviews, each placename was read both by itself for the direct case and in frames designed to elicit the oblique and ablative cases. As a result of the three-part interview, we obtained several tokens of many placenames, along a scale of more casual style in the sociolinguistic and map tasks to more formal in the word list.

The placenames in each audio file were transcribed using Praat (Boersma, 2001). Up to five of the following transcription tiers were used:

- English: one spelling for each placename was used as an index for each utterance of a given place, regardless of any particularities in individual utterances.
- Native: Pashto or Dari spelling.
- Phonetic: fairly broad transcription in the International Phonetic Alphabet (IPA).
- Language: Dari or Pashto. In general, a given task was in one language at a time. However, when working with bilingual subjects, they would occasionally explicitly remark on the pronunciation of the placename in the other language, so it was necessary to indicate the language for each placename.
- Case: for Pashto, indication of whether the particular utterance was in the direct, oblique or ablative case.

---

[1] Square brackets [] are used for allophones or sequences when no particular phonemic claims are being made; slashes // are used for phonemes.

The purpose of the phonetic transcription was to represent a human phonetic judgment that could be compared to the hypotheses of the phonetic speech recognition engine. In its documentation, Nexidia provides separate phoneme inventories in SAMPA for Dari and Pashto[2]. Sometimes Nexidia does not provide a symbol to express certain distinctions. For example, a schwa [ə] is provided for Pashto, but not Dari. Schwa is a phoneme in Pashto; however, in Dari it is a possible allophone of some short vowels. In order to facilitate experimentation with the system, in the course of phonetic transcription of a given language, we limited ourselves to the speech recognizer's phoneme inventory for that language.

3   Placename pronunciation variation

The field of toponym resolution attempts to identify which particular place, or geocode, a given place-name refers to: for example, in some contexts "London" may refer to a place in England; in others, to a place in Canada. Research in this field has primarily focused on clues in surrounding text or audio to disambiguate such placenames (Leidner, 2007; Buscaldi, 2010). To our knowledge, pronunciation variation in placenames has not yet been exploited to assist in disambiguation.

Pronunciation of placenames is well known to vary; indeed one example of this is the phenomenon known as "local pronunciation" (Forster, 1981). Some common examples from the English-speaking world include Cairo [kero], Illinois vs. Cairo [kajro], Egypt, and Houston [hawstən] Street in New York City vs. Houston [hjustən], Texas. The notion of local pronunciation is even more salient in a bilingual society; for example, French-speaking inhabitants of Montreal call their city [mɔ̃real], while English-speaking residents say [mʌntriɒl], not to mention Americans, who might say [mɑntriɔl].

In Afghanistan, Pashto and Dari are the principal languages among many other languages spoken (Farhadi, 1955; MacKenzie, 1959). Pashto and Dari-speaking communities are both located throughout the country, so it is very common for placenames to have Pashto and Dari variants, as well as variants for particular dialects of each language.

Table 1 illustrates some variation within Dari pronunciation of placenames that we encountered. This variation is not limited to placenames; in fact, each of the phonetic phenomena has been reported by Farhadi (1955), and one variant of each word may be deemed formal and the other colloquial.

Phenomenon	Place	Formal	Colloquial
/h/ dropping and compensatory lengthening	Herat هرات	[hɛrɑt]	[erat]
/ʔ/ dropping and compensatory lengthening	Qalah-ye Now قلعه نو	[qalaʔɛnaw]	[qalaɛnaw]
/a/ → [aj] / _n	Panjsher پنجشیر	[panʃer]	[pajnʃer]

Table 1: Pronunciation variation within Dari

Table 2 illustrates placename pronunciation variation within Pashto. The southwest dialect of Pashto, including Kandahar, pronounces the Pashto letter ښ as /ʂ/, while the northeast dialect, including Peshawar, Pakistan and neighboring regions of Afghanistan, pronounces it as /x/ (Miller, 2014).

Phenomenon	Place	Southwest	Northeast
/ʂ/ ~ /x/	Lashkar Gah لښکر ګاه	[laʂkargɑ]	[laxkargɑ]
/ʂ/ ~ /x/	Maydan Shar میدان ښار	[majdɑnʂar]	[majdɑnxɑr]

Table 2: Pronunciation variation within Pashto

Table 3 illustrates variation in Pashto based on case. Pashto has three cases, which may cause the pronunciation of placenames to vary. The direct case is used by default, the oblique case is used when

the placename is the object of certain prepositions and when the placename is the subject of transitive sentences in the past tense, and the ablative (also known as oblique II) is used in certain prepositional constructions meaning "from" (Penzl, 1955). Not all placenames exhibit variation based on case. Interestingly, the words that do feature a distinct oblique case take a plural ending. One interview subject suggested that in that case, the word may be interpreted as a group of people or tribe.

Place	Direct	Oblique	Ablative
Kabul كابل	[kɑbʊl]	[kɑbʊl]	[kɑbʊlə]
Bamyan باميان	[bɑmjɑn]	[bɑmjɑno]	[bɑmjɑnə]
Wardak وردک	[wardag]	[wardago]	[wardagə]

Table 3: Case variation within Pashto

Table 4 illustrates pronunciation variation between Dari and Pashto for particular places, reflecting language differences reported in Miran (1969), Penzl (1955), and elsewhere. When the native spelling used is common between the two languages, it is placed in the "Place" column; when it differs, it is placed in the "Dari" and "Pashto" columns.

Phenomenon	Place	Dari	Pashto
Dari /ɛ/ ~ Pashto /ɪ/	Helmand هلمند	[hɛlmand]	[hɪlmand]
Pashto final devoicing	Faryab فارياب	[fɒrjɒb]	[fɑrjɑp]
Dari /ɒ/ ~ Pashto /ɑ/	Kapisa كاپيسا	[kɒpisɒ]	[kɑpisɑ]
Dari /r/ ~ Pashto /ɽ/	Kunar كنر	[konar]	[kunaɽ] كنړ
Dari /q/ ~ Pashto /k/	Qalah-ye Now قلعه نو	[qalaʔɛnaw]	[kalaenaw]

Table 4: Variation between Dari and Pashto

As can be seen in Table 4, the vowel systems of Dari and Pashto differ somewhat. Dari generally employs a more rounded long *a*, which we can abstractly label /ā/, compared to Pashto. That is, Dari often uses /ɒ/ in contrast to Pashto /ɑ/. The SAMPA provided by Nexidia for each language only contains one /ā/ per language, so it is not possible to assess the system's efficacy at recognizing the rounded or unrounded variant by searching within one language; however, a method involving crosslingual search will be discussed below. In addition, future research will aim to measure the acoustic properties of the two varieties of /ā/.

With regard to consonants, Pashto has a retroflex /ɽ/, while Dari does not. In Kunar, the Pashto /ɽ/ corresponds to Dari /r/. Note, however, that when speaking Pashto as a second language, Dari speakers replace Pashto /ɽ/ with /l/ more often than /r/ (Miran, 1969). Dari preserves the Arabic voiceless uvular stop /q/, in contrast to Pashto, which generally employs /k/ in words derived from Arabic spelled with the letter ق (Penzl, 1955).

4 Assessment technique

Precision and recall are the most common measures for assessing quality in the context of audio hotspotting (Hu et al., 2012). We employ these metrics in two scenarios: dialect-agnostic and dialect-specific search. In the dialect-agnostic case, one would search for an orthographic term, for example Lashkar Gah, and calculate precision (true positives/(true positives + false positives)) based on how many of the recalled terms were in fact Lashkar Gah, and calculate recall (true positives/(true positives + false negatives)) based on how many of the actual Lashkar Gah's in the file being searched were identified. This method provides a way of evaluating the efficacy of a given system to retrieve audio of interest when one's primary concern is the place or term in question, regardless of the pronunciation that was used.

We modify the scoring method in the dialect-specific case, in which we are focused on pronunciation. Consider for example, the two common pronunciations of Lashkar Gah in Pashto: [laxkargɑ] and [laṣkargɑ]. In this case, when calculating precision, if one searches for [laxkargɑ] and [laṣkargɑ] is retrieved, it is just as wrong as if Kabul were retrieved (variable scoring, by incorporation of approaches such as Nerbonne and Heeringa (2010), will be considered in the future). For calculating recall, the universe of Lashkar Gah's is limited to those whose pronunciation matches the search term.

There is some pronunciation variation that does not necessarily represent dialect variation, and should be considered "under the radar" for the purposes of a dialect-specific search. In the example above, either of the first two vowels could be [ə] instead of [a]. For this reason, we introduce the notion of equivalence classes to enable us to give equal "correct" scores for example to both [laxkargɑ] and [ləxkargɑ] when searching for [laxkargɑ].

This scoring method provides a way of evaluating a given system's sensitivity to pronunciation differences. If a system proves adept at such a task, it can be employed in two related tasks:

- Language-specific search: find tokens of a given word uttered in a particular language
- Dialect-specific search: find tokens of a given word uttered in a particular dialect or accent

There is a large literature on language, speaker and dialect identification (Biadsy, 2011). Most of these methods are designed to emit a judgment as to language, speaker or dialect, based on a given audio sample, which might be useful in various kinds of batch processing. Another approach to accent and nativeness judgment is described by Weinberger and Kunath (2011). In this approach, audio is first reduced to a human-made phonetic transcription that is then mined for clues as to dialect and accent.

The work described here may be situated between automatic techniques based on audio and post-hoc techniques focused on transcriptions. Our method is designed for users interacting with a given audio sample; one that is likely to contain a mix of speakers, languages or dialects. Also, in contrast to statistical approaches which may appear as a "black box" to end-users, our approach allows users to iteratively and interactively develop hypotheses as to the association of specific pronunciations with languages, dialects or speakers.

5   Dialect Search

In this section, we contrast performance on dialect-specific vs. dialect-agnostic searches. Suppose in Dari we are interested in finding speakers who use the pronunciation [qalɑɛnaw] instead of [qalaʔɛnaw] for the town Qalah-ye Now قلعه نو. In this case, we are focused on the application of the phonetic process /aʔ/ → [ɑ]. The most salient aspect of this is the presence of the vowel [ɑ] rather than [a] in the second syllable. Consequently, we are unconcerned about other forms of variation we may encounter, such as variation between [q] and [k], and [ɛ] and [e]. We therefore contrast the following two equivalence classes for this experiment as shown in Table 5:

No compensatory lengthening	Compensatory lengthening
[qalaʔɛnaw]	[qalɑɛnaw]
[qalaʔenaw]	[qalɑenaw]
[kalaʔɛnaw]	[kalɑɛnaw]
[kalaʔenaw]	[kalɑenaw]

Table 5: Equivalence classes for Qalah-ye Now Experiment

When we search for a "no compensatory lengthening" pronunciation, we have a correct answer when we retrieve any one of the "no compensatory lengthening" pronunciations, and equivalently for the "compensatory lengthening" pronunciations. Table 6 provides results for precision and recall on this search above two levels of phonetic recognizer confidence:

Search Term	Confidence	Precision	Recall	True Pos.	False Pos.	False Neg.
[qalaʔɛnaw]	80	0.88	0.50	7	1	7
	60	0.80	0.57	8	2	6
[qalɑɛnaw]	80	0.33	1.00	1	2	0
	60	0.13	1.00	1	7	0

Table 6: Dialect-specific results on compensatory lengthening in Dari

As expected, recall is better with lower confidence and precision is better with higher confidence. Note that when searching for [qalaʔɛnaw], [qalɑɛnaw] is not retrieved above confidence 60. However, when searching for [qalɑɛnaw], [qalaʔɛnaw] is sometimes retrieved above that confidence level. This asymmetric performance is reflected in the higher precision values for [qalaʔɛnaw] as compared to [qalɑɛnaw].

Table 7 presents data for a dialect-agnostic search for Qalah-ye Now. For this search, we are not concerned about the particular pronunciation, so any pronunciation of the place in question will count as correct. As can be seen, this perspective causes precision to increase for [qalɑɛnaw].

Term	Confidence	Precision	Recall	True Positive	False Positive	False Negative
[qalaʔɛnaw]	80	0.88	0.47	7	1	8
	60	0.80	0.53	8	2	7
[qalɑɛnaw]	80	1.00	0.20	3	0	12
	60	0.75	0.40	6	2	9

Table 7: Dialect-agnostic results on Qalah-ye Now

Table 8 provides dialect-specific results on the diagnosis of southwest vs. northeast Pashto on the basis of the presence of [ʂ] or [x] for the Pashto letter ښ in the pronunciation of the town Lashkar Gah لښکر گاه. In the dialect-specific search, presence of [ʂ] or [x] must match between the search term and what is retrieved. The search with [x] is seen to be more precise.

Term	Confidence	Precision	Recall	True Positive	False Positive	False Negative
[laʂkargɑ]	80	0.50	0.50	1	1	1
	60	0.50	1.00	2	2	0
[laxkargɑ]	80	1.00	0.17	1	0	5
	60	0.75	0.50	3	1	3

Table 8: Dialect-specific results on /ʂ/ vs. /x/ in Pashto

Table 9 presents data for a dialect-agnostic search for *Lashkar Gah*. In this search, any pronunciation of the town will count as correct. Again, precision is seen to increase from this perspective.

Term	Confidence	Precision	Recall	True Positive	False Positive	False Negative
[laʂkargɑ]	80	1.00	0.25	2	0	6
	60	0.80	0.50	4	1	4
[laxkargɑ]	80	1.00	0.13	1	0	7
	60	1.00	0.50	4	0	4

Table 9: Dialect-agnostic results on /ʂ/ vs. /x/ in Pashto

6 Crosslingual Search

Crosslingual search is treated as a form of query expansion by Hu et al. (2012) and its efficacy as well as algorithms for its implementation in the domain of placenames are discussed by Joshi et al. (2008). We adduce crosslingual search as a tool for assessing language-specific search. For example, if we search for Kabul using the Pashto engine, to what extent will we retrieve Pashto utterances of that place as opposed to Dari utterances, and vice versa? If the Pashto engine is good at picking up Pashto to the exclusion of Dari utterances of a placename, it may be an effective tool for language-specific search.

We performed a set of experiments to assess this capability. First, we performed a search that was agnostic with respect to language and dialect. This means that in a search for Kabul in Pashto, we give credit for both Pashto and Dari tokens of Kabul that are retrieved, regardless of their particular pronunciations. Next, we performed language-specific searches in both Dari and Pashto. When searching in a given language, we only give credit for retrievals in that language. Note that when we performed language-specific search, we were dialect-agnostic. That is, we gave credit for a retrieval provided it was in the language being searched for, regardless of the particular pronunciation used.

The first term used for both language-agnostic and language-specific search was IPA [kɑbʊl]. Note that due to details of the Nexidia engine, the actual SAMPA strings used were [k A: b O l] for Dari and [k A b u l] for Pashto. The symbols for /ā/ and /ʊ/ in each language are arbitrarily different as indicated in Table 10. While IPA symbols (and their SAMPA equivalents) are theoretically absolute values in acoustic or articulatory space, in practice, they often adhere to arbitrary conventions for transcription of a particular language.

Language	Orthographic symbol	IPA	SAMPA
Dari	١	ɑ, ɒ	A:
Pashto	١	ɑ	A
Dari	̊	ʊ	O
Pashto	̍	ʊ	u

Table 10: Differences in phoneme symbols used for Dari and Pashto

Pashto exhibits pronunciation variation between [kɑbʊl] and [kɑbəl]. Table 11 compares performance on language-agnostic search performed for Kabul in each language.

Language	Search	Confidence	Precision	Recall	True Pos.	False Pos.	False Neg.
Dari	[kɑbʊl]	60	0.75	0.61	41	14	26
Pashto	[kɑbʊl]	60	0.80	0.24	16	4	51
	[kɑbəl]	60	0.82	0.21	14	3	53

Table 11: Language and dialect-agnostic search

Table 12 compares performance on language-specific search. Note that this search was still dialect-agnostic, so credit was given as long as the token was in the searched-for language, regardless of its pronunciation.

Language	Search	Conf.	Prec.	Recall	True Pos.	False Pos.	False Neg.
Dari	[kɑbʊl]	60	0.29	0.42	16	39	22
Pashto	[kɑbʊl]	60	0.70	0.37	14	6	24
	[kɑbəl]	60	0.82	0.37	14	3	24

Table 12: Language-specific search

As we can see from these results, the Dari engine has better precision and recall on the language-agnostic search, in contrast to the Pashto engine, whose recall is better on language-specific search. This can be interpreted as follows: the Dari engine is more versatile and can pick up Pashto, whereas the Pashto engine is more specific to Pashto and does not pick up Dari as well.

7 Conclusion

We have achieved some success searching for language and dialect-specific pronunciations using the Audio Gazetteer tool. A future challenge will be to identify dialect-specific toponyms automatically from a gazetteer. Our results are encouraging for the exploitation of pronunciation variation in toponym resolution and perhaps speaker identification. While dialect-specific results are often not as precise as searches that are agnostic as to language or dialect, in effect because we are "raising the bar" for what

is correct, more data and more dialect and language-specific phenomena need to be collected and processed through the system in order to establish its capabilities more clearly.

References

Fadi Biadsy. 2011. *Automatic dialect and accent recognition and its application to speech recognition.* Ph.D. dissertation, Columbia University.

Paul Boersma. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10):341-345.

Davide Buscaldi. 2010. *Toponym disambiguation in information retrieval.* Ph.D. dissertation. Universidad Politécnica de Valencia.

Abd-ul-Ghafur Farhadi. 1955. *Le persan parlé en Afghanistan.* Klincksieck, Paris.

Klaus Forster. 1981. *A pronouncing dictionary of English place-names including standard local and archaic variants.* Routledge, London.

Michelle Annette Minnick Fox. 2006. *Usage-based effects in Latin American Spanish syllable-final /s/ deletion.* Ph.D. dissertation, University of Pennsylvania.

Marsal Gavalda and Jeff Schlueter. 2010. "The truth is out there": Using advanced speech analytics to learn why customers call help-line desks and how effectively they are being served by the call center agent. In Amy Neustein, editor, *Advances in speech recognition: Mobile environments, call centers and clinics,* pages 221-243. Springer, New York.

Qian Hu, Fred J. Goodman, Stanley M. Boykin, Randall K. Fish, Warren R. Greiff, Stephen R. Jones, and Stephen R. Moore. 2012. Automatic detection, indexing, and retrieval of multiple attributes from cross-lingual multimedia data. In M. T. Maybury, editor, *Multimedia information extraction,* pages 221-233. Wiley, Hoboken.

Tanuja Joshi, Joseph Joy, Tobias Kellner, Udayan Khurana, A Kumaran, A. and Vibhuti Sengar. 2008. Crosslingual Location Search. *SIGIR '08.*

Jochen Lothar Leidner. 2007. *Toponym resolution in text.* Ph.D. dissertation, University of Edinburgh.

D. N. MacKenzie. 1959. A Standard Pashto. *Bulletin of the School of Oriental and African Studies* 22(1/3):231-235.

Corey Miller, Rachel Strong, Evan Jones and Mark Vinson. 2013. Reflections on Dari linguistic identity through toponyms. In Rudolf Muhr et al., editors, *Exploring linguistic standards in non-dominant varieties of pluricentric languages,* pages 319-330. Peter Lang, Vienna.

Corey Miller. 2014. Pashto Dialects. In Anne Boyle David, *Descriptive grammar of Pashto and its dialects,* pages 32-44. Mouton De Gruyter, Berlin.

Mohammad Alam Miran. 1969. *Major problems of Dari speakers in mastering Pashto morphology.* M.A. Thesis, UT Austin.

John Nerbonne and Wilbert Heeringa. 2010. Measuring dialect differences. In J. E. Schmidt and P. Auer, editors, *Language and space: an international handbook of linguistic variation, volume 1, theories and methods.* Mouton de Gruyter, Berlin.

MITRE. 2012. The MITRE Corporation Annual Report.

Herbert Penzl. 1955. *A Grammar of Pashto.* American Council of Learned Societies, Washington.

Raytheon. 2012. BBN Broadcast Monitoring System. Retrieved from http://bbn.com/resources/pdf/bms.pdf

Steven H. Weinberger and Stephen A. Kunath. 2011. The speech accent archive: towards a typology of English accents. In J. Newman, H. Baayen, H., and S. Rice, editors, *Corpus-based studies in language use, language learning and language documentation,* pages 265-281. Rodopi, Amsterdam.

J. C. Wells. 1997. SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore and R. Winski, editors, *Handbook of standards and resources for spoken language systems.* Mouton de Gruyter, Berlin.

Mirjam Wester, Judith M. Kessens, Catia Cucchiarini and Helmer Strik. 2001. Obtaining phonetic transcriptions: A comparison between expert listeners and a continuous speech recognizer. *Language and Speech* 44(3): 377-403.