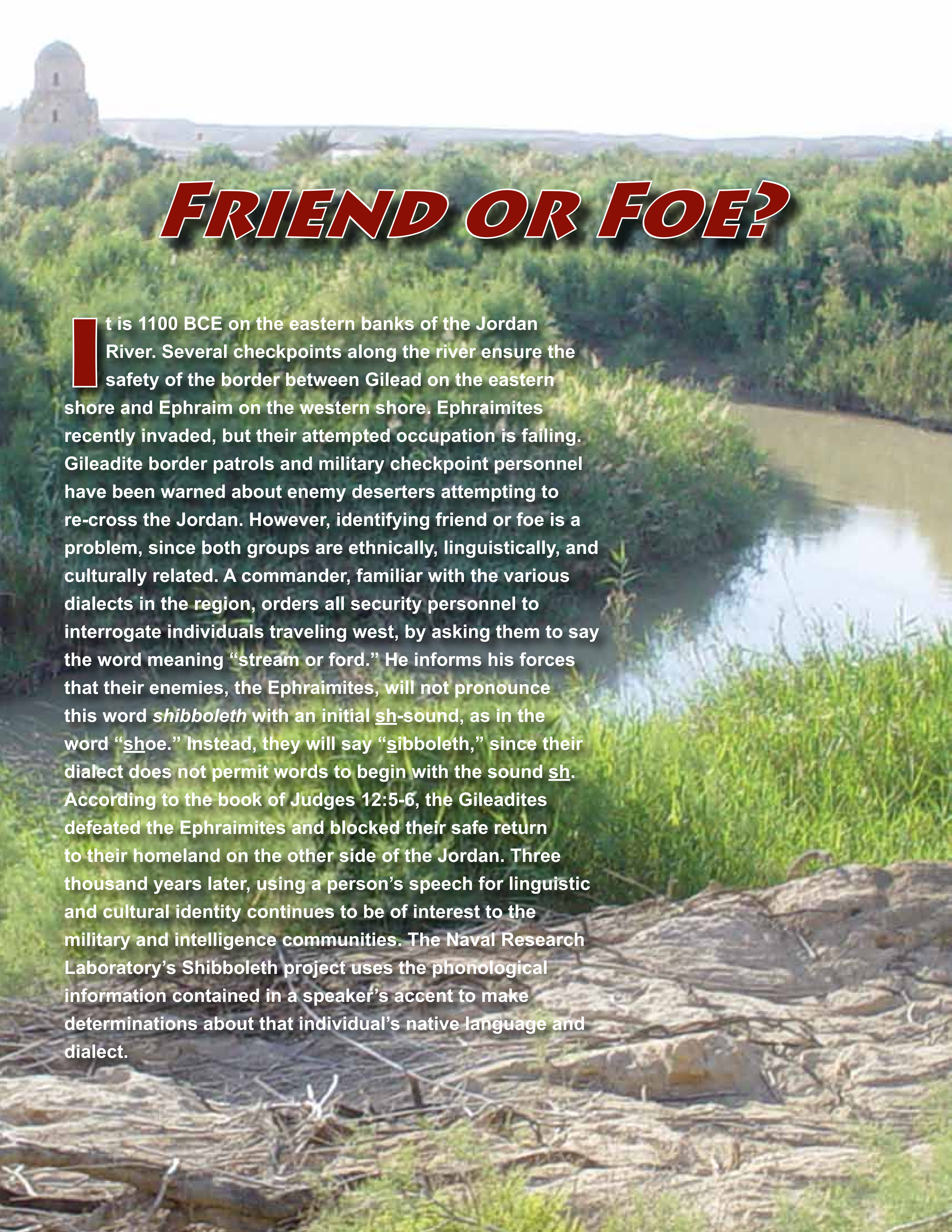


Friend or Foe?

It is 1100 BCE on the eastern banks of the Jordan River. Several checkpoints along the river ensure the safety of the border between Gilead on the eastern shore and Ephraim on the western shore. Ephraimites recently invaded, but their attempted occupation is failing. Gileadite border patrols and military checkpoint personnel have been warned about enemy deserters attempting to re-cross the Jordan. However, identifying friend or foe is a problem, since both groups are ethnically, linguistically, and culturally related. A commander, familiar with the various dialects in the region, orders all security personnel to interrogate individuals traveling west, by asking them to say the word meaning "stream or ford." He informs his forces that their enemies, the Ephraimites, will not pronounce this word *shibboleth* with an initial sh-sound, as in the word "shoe." Instead, they will say "sibboleth," since their dialect does not permit words to begin with the sound sh. According to the book of Judges 12:5-6, the Gileadites defeated the Ephraimites and blocked their safe return to their homeland on the other side of the Jordan. Three thousand years later, using a person's speech for linguistic and cultural identity continues to be of interest to the military and intelligence communities. The Naval Research Laboratory's Shibboleth project uses the phonological information contained in a speaker's accent to make determinations about that individual's native language and dialect.

How Do *You* Say It?
Shibboleth: Phonological Analysis of Non-Native Speakers of English

W.K. Frost, D.J. Perzanowski, and R.C. Page
Information Technology Division

To assist in the war on global terrorism, the Naval Research Laboratory's Shibboleth project is investigating the speech of non-native speakers of English. At U.S. military checkpoints around the world, warfighters typically speak in English when they interrogate individuals wishing to cross the checkpoints. The respondents, non-native English speakers, reply in English. By studying the speech of these non-native speakers, linguists at NRL are determining the phonological rules that characterize or distinguish accents of non-native English speakers. With this information, we can determine the native language of the speakers using rule-based phonological patterns. This helps to identify potential threats when, for example, an individual is trying to hide something or avoid identification by lying about himself or herself. Knowing the phonological identity or correctly identifying the accent of an individual can aid the warfighter in determining whether persons of interest are telling the truth or lying about their identities.

LANGUAGE AS AN INDICATOR

From Biblical times to World War II, military and intelligence-gathering individuals have been using dialect variation and accents to isolate what are called phonological "shibboleths" to identify individuals' backgrounds, nationalities, and cultures, and in so doing distinguish friend from foe. For example, U.S. forces stationed in northern Europe used one shibboleth to distinguish friendly Dutch-speaking forces and citizens from the potential German-speaking enemy. To the untrained American listener, Dutch and German speakers sound similar, as do their accents when these individuals are speaking English. If a U.S. soldier was doubtful about an individual's identity or background, he knew that he could easily distinguish between the two nationalities by asking the individual to pronounce the name of the Dutch seacoast town, Schevenigen. A German, untrained in the Dutch language, would have some difficulty replicating the complex word-initial consonant cluster of "**Sx**evenigen," instead pronouncing it as "**Sh**evenigen." Today, in our post-9/11 world, quick identification of friends and foes, terrorists and non-terrorists, has become a crucial aspect of national defense and a main component of the global war on terrorism.

To aid in the identification of potential threats, various biometric devices are used in conjunction with metal and explosive detectors. Retinal scans and fingerprinting, for example, are used, but these methods take time and resources that may not be readily available at a busy or remote border station or checkpoint. To supplement and enhance these techniques, the Naval Research Laboratory's Shibboleth project seeks to enable phonological analysis of individuals' speech to determine a person's linguistic background, by comparing and contrasting known phonological variations in speech and accents with those of suspect persons. While individuals might hide behind the guise of another nationality, using a forged passport, for example, people have a difficult time hiding their linguistic identity. People may attempt to train themselves to speak differently, by mimicking a different dialect or accent, but phonological analyses of their speech patterns can detect the disguise or inconsistency. This type of analysis can be used today in countries like Afghanistan, where six different native languages are dominant in its 34 provinces (Fig. 1). Even for a linguistically trained warfighter who may be familiar with one or two of those languages, this is a daunting situation.

THE PHONOLOGICAL BASES OF LANGUAGE

Speech is the result of various muscle movements in the oral cavity that incorporates the nasal passage, mouth, and lips, and extends to the larynx or pharynx (Fig. 2). These movements, coupled with air expelled from the lungs and modulated by the vocal cords, create differences in sound and produce changes in air pressure. These differences can be seen in a spectrogram (Fig. 3) that records the amount of air pressure produced over time. The more vividly colored bands

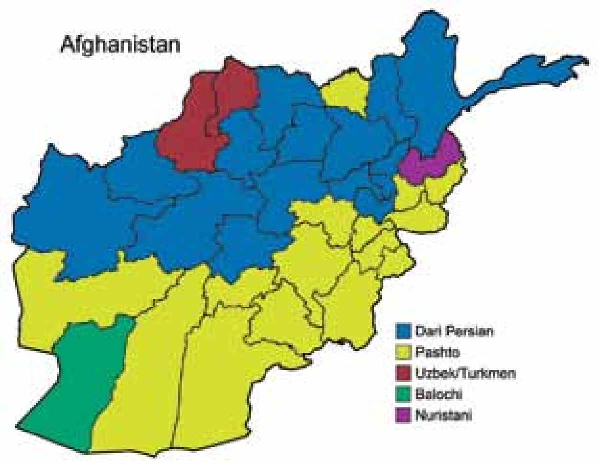


FIGURE 1
Majority languages spoken in the provinces of Afghanistan.

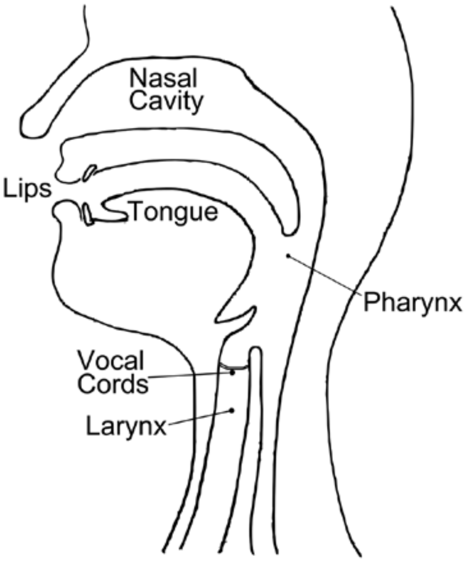


FIGURE 2
Cutaway diagram of the human vocal tract.

(pink and blue) in the spectrogram, for example, exhibit a concentration of air pressure measured in hertz (Hz). Typical human speech ranges from 300 to 3500 Hz. Because varying air pressure and altering the configuration of the various speech apparatuses mentioned above causes all human speech sounds, spectrograms offer an excellent depiction of what speech sounds look like when visually mapped to a representation of what is happening inside the human anatomy. A gloss in Fig. 3 indicates where the various sounds or "phones" of the utterance "Show me your I.D." occur, and the spectrogram offers a visual representation of the amount of acoustic energy or air pressure measured at the various points along the speech continuum.

This provides information about various acoustic features that characterize speech sounds. For example, the difference between the initial sound or "phone" in *pat* and *bat* is what is known as "voicing." In *pat*, the initial sound is called "voiceless" because no air is being forced out of the lungs and modulated by the vocal cords. On the other hand, in *bat*, the initial sound is "voiced" because air is being expelled from the lungs causing the vocal cords to vibrate. Other distinctive phonetic features can be used to identify speech sounds. Typically, these features are constrained to a binary representation so that the feature [+/– voice] is realized as [+voice] for the phone **b** and [–voice] for the phone **p**.

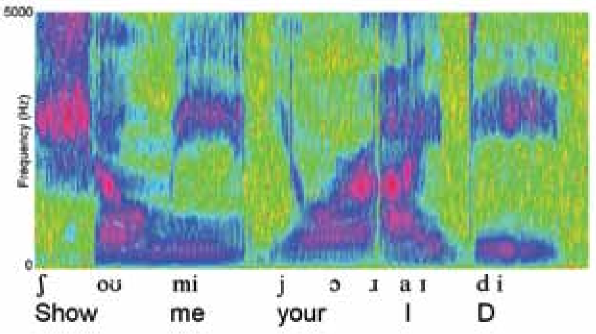


FIGURE 3
Spectrogram of the phrase "Show me your I.D."

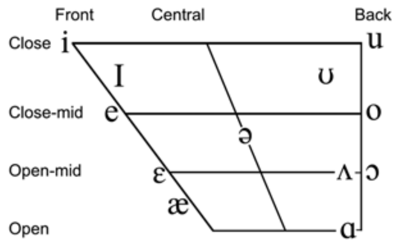
Coupled with the acoustic depiction of speech sounds is the classification of speech sounds by manner of articulation. This set of classifiers captures how the musculature is manipulated and how various articulators, such as the tongue and teeth, are used to produce a particular sound. For example, we say that both **b** and **p** are bilabial, since they employ the use of the lips in producing the sounds. Thus, these two phones employ the feature of [+/– bilabial]. Both **b** and **p** are [+bilabial] and with the acoustic feature discussed above, **b** = [+voice, +bilabial] and **p** = [–voice, +bilabial]. Using such features, we distinguish these two phones, but we also distinguish them from other phones, such as **d** = [+voice, –bilabial] and **t** = [–voice, –bilabial]. Using this system of binary classification, we can distinguish the various speech sounds of the world's languages. The International Phonetic Association (IPA), founded in Paris in 1886, continues to have as its aim the creation and maintenance of a standard of phonetic representation for all of the sounds of the world's languages. (See Fig. 4 for a representation of an IPA chart depicting the sounds of American English.)

However, human speech does not consist of individual phones uttered, but of sequences of sounds in various combinations. Phonology is the science that constructs rules that characterize how the various phones of the world's languages are combined. Languages exhibit rules that either permit or restrict

Consonants

	Bilabial	Labiodental	Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	ʔ
Nasal	m			n			ŋ	
Flap				ɾ				
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant				ɹ		j		
Lateral Approximant				l				

Vowels



Other Symbols

ʍ	Voiceless labial-velar fricative
w	Voiced labial-velar approximant
dʒ	Voiced papato-alveolar affricate
tʃ	Voiceless palato-alveolar affricate

FIGURE 4
IPA charts of American English consonants and vowels.

the combination of certain sounds in that language. We have seen examples of some restrictions in historical examples above. The phonologist determines what those permissions and constraints are and characterizes them for various human languages. By doing so, the phonologist is both depicting what is permissible in a particular language and predicting what is impermissible. A native speaker of a language intuits these rules, and non-native speakers either have difficulty emulating those speech patterns or never learn them. This gives rise to the tell-tale "shibboleths" in their speech.

SHIBBOLETH: A TOOL FOR ACCENT IDENTIFICATION

The process of automatically determining a speaker's first language or dialect from the accent perceived in a second language has two main components: rule learning and accent classification. Shibboleth, our prototypical tool for phonological analysis, first learns phonological rules from phonetic transcriptions for each language or dialect to which it is introduced, and compiles these rules into rule sets. These rule sets serve as a constantly evolving repository of information describing the phonological space made up of any dialect in any language. This mechanism enables users in the field to train Shibboleth on previously unmapped languages or dialects and lessens reliance on analyses from trained linguists. Thus far, Shibboleth has used the data in the George Mason University Speech Accent Archive,[1] an archive of speakers from around the world, to create and to modify its rule sets.

Training data, used to initialize the learning of these rules sets, shows the phonological representation of the influences of a speaker's native language or dialect when the sample is compared to the "standard" transcription. This can be used to observe differences from a speaker's native language to a second language learned as an adolescent or adult, or to note when a speaker is speaking a non-standard dialect of a language, such as speakers of one of the regional U.S. dialects, Southern American English. The process of creating rule sets for individual dialects within a language is identical to that for creating rule sets for another language. With a dialect, however, the phonological differences will generally be more subtle and harder to detect.

The Shibboleth program distinguishes differences by tracking every time a sound or phone varies between a speaker of an unidentified dialect and a generic American speaker. It catalogues both the change itself, the number of times the change occurred, and the environment in which the change occurred. For example, it notes when a speaker switches the **l** and the **r** sounds in a word, such as saying "lice" for the word "rice." It also notes if the speaker did so consistently, or only at the beginning of words. It also notes if this exchange happened every time at the beginning of the word that should have started with an **r**, or if it only occurred once.

An example of a common phone change in American English is the merging of **ɛ** and **ɪ** in some dialects, such as Southern American English. This change would cause the words "pen" and "pin" to become homophonous, or sound identical, for speakers of that dialect. This one relatively simple phone change would give rise to 320 rules, creating a combinatorial explosion that the program must process over the course of an utterance. Asking even a trained linguist to manage such a large set of rules is awkward, cumbersome, time-consuming and error-prone. Therefore, Shibboleth is being designed to relieve individuals, such as warfighters, of such a burden.

Depending upon the specificity of the rule (i.e., if the rule is based on one phonetic feature, such as voicing) or its generality (i.e., if it is based on several features, such as voicing and a particular place of articulation), the amount of evidence for the rule from the speakers, and the probability of the rule, given the sample, each rule in the rule set is assigned a weight using a Bayesian average as an additive smoothing algorithm to adjust the score. The use of the Bayesian average adjusts rule scores for rules whose environments occur either much more frequently than the average environment (e.g., the combination **ing** at the end of words in English) or much less frequently than the average environment (e.g., the cluster **thr** in English).

After analyzing various languages, dialects, or language families, Shibboleth stores the resultant rules in a database, and the system moves on to accent classification. Speech samples in which the speaker's native language or dialect is unknown or in doubt, such as those of a speaker at a border checkpoint, are compared to the existing rule sets to find the language or dialect that is the most likely match. The strength of the match to the rule set is determined by the deviation from the mean score of the proposed matched rule set, and provides the user with a certainty rating of "high," "moderate," or "low." In addition to pinpointing native languages or dialects, Shibboleth's rule sets can also be a useful tool in recognizing lesser known or infrequently occurring phonological phenomena across languages or in gathering data on how dialects might change due to an individual subject's demographic information and background.

LANGUAGE AND DIALECT RESULTS

We have tested Shibboleth on a preliminary database of Italian and German rule sets, as well as on four major American English dialects. The results have been largely positive, with Shibboleth correctly determining individual accents at the language level in 82% of the speakers examined, and at the dialect level for 77% of the speakers.

The language differentiation task contained 44 speakers, comprised of 15 native Italian speakers, 15 native German speakers, and 14 native American English speakers, all of whom spoke in English. From a 69-word speech sample, Shibboleth correctly identified 9 of the Italian speakers, 14 of the German speakers, and 13 of the American English speakers, as shown in Fig. 5. A speaker who scored between 0 and 0.1 is considered to have unaccented speech, where 0 is completely unaccented American English. A speaker with a score in any foreign language higher than 0.1 is considered to be a non-native speaker of American English. If multiple scores for a speaker were higher than 0.1, the highest score is considered to be the likely native language of the speaker. It should, therefore, be noted that Shibboleth correctly identified the speakers better than chance in all cases.

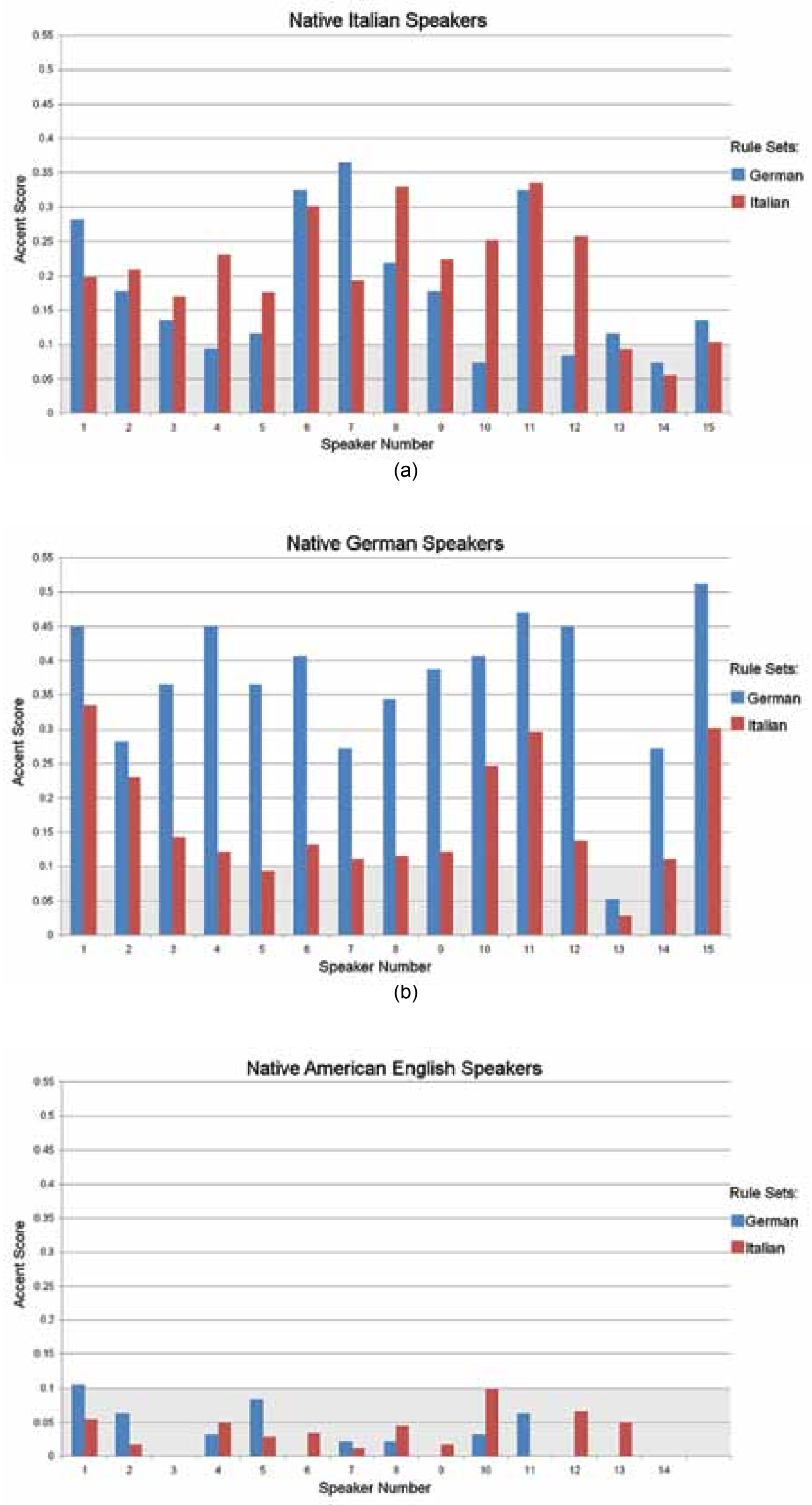
As mentioned previously, dialect comparisons are more difficult, as the divergence from a standardized sample will be smaller. The dialect differentiation task performed with Shibboleth compared Northern, Southern, Midland, and Western American English dialects, and the speaker sample was comprised of seven Northern dialect speakers, seven Southern dialect speakers, four Midland dialect speakers, and four Western dialect speakers. From the same sample, Shibboleth correctly identified with some level of certainty five of the Northern speakers, five of the Southern speakers, three of the Midland speakers, and all four of the Western speakers. Because of the small sample size, we do not wish to comment on the accuracy of these identifications. Furthermore, the certainty on many of these placements was low, indicating the need for additional research to obtain more classification accuracy.

SUMMARY

The Shibboleth program recognizes the likely native language and dialect of an individual by analyzing the person's accent in a second language or dialect, here, American English. It can be trained to recognize the accents of previously unsampled languages and dialects in the field, making it an adaptable tool for the warfighter or border guard who may not have access to other databases or have a specialized language background to make complicated linguistic distinctions of the type outlined here. Initial tests of the system have given positive results, although certainty ratings of the classifications, particularly on dialect variation, need to be strengthened.

ACKNOWLEDGMENTS

The authors thank Steven Weinberger from George Mason University, Stephen Kunath from Georgetown



(a)

(b)

(c)

FIGURE 5

Accent scores for (a) native Italian speakers matched against German and Italian rule sets; (b) native German speakers matched against German and Italian rule sets; and (c) native American English speakers matched against German and Italian rule sets.

University, and student interns Danielle Schrimmer and Jodi Lamm.

[Sponsored by the NRL Base Program (CNR funded)]

Reference

[1] S. Weinberger, *The Speech Accent Archive*, George Mason University (2012). Retrieved from http://accent.gmu.edu.

THE AUTHORS

WENDE FROST received a master's degree in computer science in 2010 with a focus on intelligent systems and autonomous agents from Arizona State University, preceded by an M.A. in linguistics in 2007, also from Arizona State University. She obtained a B.S. in computer information systems from Northern Arizona University in 2003. She has worked at the Naval Research Laboratory in the Interactive Systems Section since 2009 following internships in the Intelligent Systems Section beginning in 2008. While working at NRL, her research interests have included machine classification of spoken language, contextual interaction, and representations of common ground with intelligent agents. She is a member of the Association for Computational Linguistics (ACL), the ACL special interest group on discourse and dialogue (SIGDIAL), and the Association for the Advancement of Artificial Intelligence.

DENNIS PERZANOWSKI earned his Ph.D. degree in linguistics from New York University in 1982 with an emphasis on theoretical syntax and semantics. He received an M.A. from New York University in linguistics in 1976 and an M.A. in English literature from Temple University in 1968. He also obtained a B.A. in English literature from La Salle University in 1965. Upon graduation from Temple University, he taught English composition at Miami Dade College. While working toward his degrees at NYU, he taught English composition at Fordham University Lincoln Center, where he was also an assistant academic dean. After receiving his Ph.D. degree, he taught English composition and linguistics at Utica College of Syracuse University. In 1984, he came to work in the Natural Language Section of the Naval Research Laboratory and has served as the section head of the renamed Interactive Systems Section since 2007. His research interests at NRL extend from human-computer/robot interfaces to natural language processing and phonological analysis. He is a member of the Association for the Advancement of Artificial Intelligence, the Association for Computational Linguistics, and the Linguistic Society of America.

ROBERT PAGE is a computer scientist for the Immersive Simulation Section in the Information Management and Decision Architectures Branch. Rob joined the Information Technology Division of NRL in 1996. He has worked in a number of research areas for the division. He has developed software to perform machine classification of spoken language, developed several generations of virtual reality software, and written software to simulate shipboard tactical displays. He received his B.A. degree in computer science from Rutgers University in 1987 and an M.S. degree in computer science from George Washington University in 1996. He is a member of the Association for Computing Machinery's special interest groups for graphics (SIGGRAPH) and computer-human interaction (SIGCHI). Rob began working at NRL in 1990 as a contractor for the Physical Acoustics Branch.