SBE Grand Challenge:
Understanding the complexity and variability of spoken and signed languages

Submitted to the National Science Foundation

Authors:

Natasha Warner	University of Arizona
Jennifer Cole	University of Illinois
José I. Hualde	University of Illinois
John Kingston	University of Massachusetts
Amalia Arvaniti	University of California San Diego
Mary Beckman	Ohio State University
Janet Pierrehumbert	Northwestern University
Douglas H. Whalen	Haskins Laboratories

ABSTRACT

How do humans (learn to) manage the complexity and variability of speech in everyday language, or equally, the complexity and variability of sign in signed languages? Complexity pervades all aspects of the spoken and signed forms used in the world's languages. Variability arises from the interplay of cognitive and physiological factors governing the production and perception of speech and sign, biological and cognitive factors governing human social behavior, and cultural factors governing linguistic communication. Complexity and variability demand an interdisciplinary approach to investigating spoken and signed utterances from a broad spectrum of individuals, communities, social functions and settings. We propose to advance the investigation of spoken and signed languages by (1) creating new interdisciplinary learning and training programs that bring talented students and scholars from all relevant disciplines together in sustained collaborative learning and research interactions, (2) creating new, large speech and sign databases that represent the full variety of styles, speakers/signers, dialects, and languages (3) advancing capabilities for intelligent datacapture (e.g., multichannel high speed AV and sensor networks), signal processing, and machine learning. This proposal builds on the successful approach of laboratory phonology, using research methods and tools from across the social and behavioral sciences, computer science and engineering.

In response to the NSF-SBE directorate's call for framing innovative research, helping to determine the direction of science for the next decade and beyond, we propose the following "grand challenge" question:

- How do humans (learn to) manage the complexity and variability of speech in everyday language?

Language is a central aspect of human cognition, an essential vehicle of social interaction, and a defining hallmark of human development. Complexity characterizes all aspects of language, including the speech forms that signal language. Enormous variability arises from the interplay of cognitive and physiological factors governing the production and perception of utterances, and from the interplay of cognitive and cultural factors governing their use in social interaction. These same factors make children's acquisition of adult community norms a highly complex and variable process. This variability is demonstrated by the fact that the same speech signal is interpreted differently by perceivers depending on their individual experience with language, on the social and linguistic context, on available visual cues, and on their inferences regarding the speaker. At the same time, objectively very different signals may be interpreted as identical depending on these factors. This complexity makes the investigation of speech a necessarily interdisciplinary endeavor, and requires the examination of data from broad spectrum of individuals, communities, social functions and settings.

A substantial body of research shows that the manual signs of signed language are analogous to the sound component of spoken languages and equally complex and variable, which leads us to a more general formulation of the grand challenge question: *How do humans (learn to) manage the complexity and variability of speech and sign in everyday language?*

Laboratory phonology is a research enterprise concerned with the nature and use of spoken and signed language which has developed over the past 25 years in response to the challenges presented by these characteristics of speech and sign. The challenges are technical, empirical and theoretical. Using research methods and tools from social, behavioral and computer sciences and engineering, tremendous progress has been achieved in understanding the nature of spoken and signed language as complex and variable phenomena. The success and future promise of the laboratory phonology approach have two sources. First, laboratory phonology research has become increasingly bottom-up, recognizing that a broad empirical foundation is crucial for new discoveries. Second, it has necessarily become more encompassing in its methods and theoretical underpinnings, drawn from many disciplines. Two capabilities need to be developed more extensively in the coming decades if we are to build on our prior success and address the "grand challenge" question fully:

Capability for dramatically increased interdisciplinary collaboration. The evidence from recent laboratory phonology conferences shows that speech and sign can only be understood by collaboration across disciplines. To achieve this goal, we propose the creation of capabilities in interdisciplinary learning and training programs that bring

talented students and scholars from all relevant disciplines together in sustained collaborative learning and research interactions. These interactions may include conferences and workshops, but the focus should be on training and research grants led by interdisciplinary teams. Researchers from the laboratory phonology community and allied disciplines should establish broad directions that would serve as the scientific and logistical umbrella for the development of these training and research grants.

Capability for the creation of new corpora, and the tools and technologies for data capture and annotation. If we are to achieve a thorough understanding of how speech and sign are used and learned in everyday interactions, researchers must have access to speech and sign databases that represent a wider variety of styles, speakers/signers, dialects, and languages than is currently available. Corpus resources must be expanded beyond existing speech databases, many of which consist of normative speech recorded under laboratory conditions, to include new larger databases that represent, e.g., speech produced by people with hearing loss, with limited proficiency in a second language, from different dialects, of different ages, in noisy environments, or in any of the other situations that reveal the full variability of linguistic communication in real life. We propose the creation, over the next ten years, of many more large spoken and signed language corpora from a variety of languages. Technologies for data capture and advances in machine learning and signal recognition will be applied and developed in tandem to meet the engineering challenges of creating and processing such large databases. The goal here is to produce substantial labeled corpora that are widely shared by many researchers and laboratories and a common set of sophisticated statistical, computational and engineering tools for their analysis. Stimulating and guiding the development of these corpora and the tools for their analysis would be one of the principal responsibilities of the group that devises broad directions for the field.

These capabilities are the scientific strategies that laboratory phonology has already begun to use, but far more needs to be done to make them truly effective means of handling the full complexity and variety of speech and sign.

II. Context

A number of recent findings have challenged traditional views regarding the mental encoding of phonology as a symbolic system to the point that the phonological edifice must be built anew, on interdisciplinary foundations. One such finding is that probabilistic processing is pervasive in speaking and listening. Another is that linguistic and non-linguistic information is communicated through multiple channels at once; the manual and facial gestures that accompany speech are interpreted together with the information present in the acoustic signal. Furthermore, speakers produce reduced or disfluent speech in the regular course of speaking, but reductions and disfluencies normally go unnoticed and do not disturb comprehension. To complicate matters, listeners simultaneously extract linguistic meaning and nonlinguistic information (regarding the social and individual characteristics of speakers and their emotions) from the signal. The same considerations apply to manually signed language. All of this forces us to question how phonology is encoded in the mind/brain of an individual, to what

extent members of the same linguistic community can be said to possess the same phonological system, and whether phonological representations used in production and perception are equivalent. The answers can only come from synergies involving linguists, other cognitive and social scientists, computer and information scientists and engineers.

Work on these questions and others is presently the central research focus in the laboratories of this white paper's authors, as well as in laboratories at USC, UCLA, Berkeley, Chicago, Minnesota, Michigan, Indiana, and NYU.

III. Range of disciplines that contribute

In recent years we have witnessed a dramatic increase in the fruitful collaboration among linguists interested in the study of speech and manual sign systems, and researchers in an increasingly large number of other fields, including psychology, cognitive science, speech and hearing science, electrical and computing engineering, and neurolinguistics. This is evident from a consideration of the departmental affiliations of the participants at the biennial Conference on Laboratory Phonology (e.g., Warren & Hay 2010; Fougeron et al., 2010; Cole & Hualde 2007) and from the diversity of academic fields where work on laboratory phonology is cited. The field is ready to include intellectual interaction with even more disciplines. In particular, the role of speech and sign communication in human interaction invites a fruitful exchange of ideas and methods with the fields of communications, sociology and anthropology. The very broad research program envisioned here also has obvious applications to the development of successful human-machine interactions using language, to remediation of hearing and speaking disorders, to college language teaching and the acquisition of languages by adults, which is the traditional domain of applied linguistics, and to K-12 education, to name just a few.

IV. Implications for future research

In any scientific field, there are questions that most researchers feel one is "not supposed to" ask, because they reveal that the goals of the field are not understood. Laboratory phonology has witnessed an explosion in the range and scope of askable questions over the past 20 years, due to developments in methods for investigating speech and signed language, and to the cross-fertilization of theories from the contributing disciplines. The future science of spoken and signed language will ask how the speech and sign systems of human language are shaped by the physical and psychological constraints imposed by production and perception, how the brain acquires and processes linguistic information and how its acquisition and processing relates to information processing in such other cognitive domains as vision and music, how language systems emerge from the varied behavior of individuals, and how linguistic expression in speech and sign is used to communicate social identity.

The biological, neural and social bases of speech and sign are now discoverable, given the increasing availability of data derived from imaging and tracking the function of the brain as language is produced, perceived and learned, from mining very large audio and video archives for speech and sign data, from tracking the gestures of the vocal tract and hands in speech and sign production, and from computational modeling of information signaling in speech and sign. Changes in theory also advance the scientific study of

speech and signed language. Theories of memory, pattern recognition and generalization in cognitive psychology are key to understanding how the rich detail of acoustic and visual signals gives rise to abstract representations of linguistic form in the mental encoding. Theories of embodied cognition are relevant to understanding how the perception and production of speech or sign are related. Theories of the behavior of complex dynamical systems provide insight into the neural basis of speech motor behavior and of the interaction between the hierarchically layered structures of phonological form.

Many, if not most of these new methods and theories were first developed in other disciplines with distinct but overlapping goals from laboratory phonology, and required further development for application to spoken and signed language. But the borrowing of methods and ideas flows in both directions. New data on speech and sign are valuable in enriching theories constructed for the goals of other disciplines as well. For example, information about the transmission of pronunciation change across members of a speech community provides evidence for the structure of social networks.

We expect a substantial benefit to arise for scientific knowledge across all the disciplines that contribute to the study of speech and sign. One impact comes from the adoption of new paradigms for collecting speech data in ecologically valid fields of everyday language, opening up new and globally expanded spheres of language use to scientific inquiry across disciplines.

Collecting speech data on a global scale to achieve diversity in form and function requires advances in computer and information science, and new technologies for capturing, coding and measuring speech and sign. Impact from these developments arises in applications to data mining and information processing in domains other than speech.

Further scientific impact is in the expansion and creation of theories of human cognitive behavior to address speech and sign perception in variable situations. For instance, current models of the perception of foreign speech explain how language learners form and maintain speech sound categories, but these models do not yet encompass how learners perceive spontaneous speech reduction, how they cope with background noise, or whether they make use of the statistical properties of sound sequences in the target language. Theories of phonology or phonetics do not yet necessarily interface with theories of second language pedagogy or neurolinguistics, and neither were they designed to model the degree of variability of spontaneous speech (even by native speakers).

The expansion of interdisciplinary training proposed here will also have an impact in tangible benefits for language teaching, and for diagnosis and interventions in speech and hearing with clinical populations.

Most broadly, building a science of how humans communicate with speech and sign in a wide variety of real-world situations would lead to benefits for the people who communicate in all those situations. Laboratory phonology serves as an obvious basic-

science point of organization for such an integrated approach to all the dimensions of real-world variability in speech and sign communication.

References

Cole, Jennifer and Hualde, José I. 2007. *Laboratory Phonology 9.* Berlin: Mouton de Gruyter.

Fougeron, Cécile, Kühnert, Barbara, D'Imperio, Mariapaola, and Valée, Nathalie. 2010. *Laboratory Phonology 10.* Berlin: Mouton de Gruyter.

Hay, Jen and Warren, Paul. 2010. *Laboratory Phonology* vol. 1.1. Special issue: Papers from the Eleventh Conference on Laboratory Phonology.