Testing hybrid models of phonological representation with speech corpora

The objective of the project

The goal of this project is to investigate the need for so-called hybrid models of storage of speech sounds in the mind. It will look for evidence in speech of effects of storage of both 'abstractions' (i.e. essential characteristics of sounds needed to make distinctions between words) and so-called 'exemplars' (i.e. memory traces of words containing much more detailed information about the sounds of words). As data source, speech corpora, i.e. large collections of recordings of conversational speech synchronized with their transcripts will be used.

Aiming to find evidence for both abstractions and exemplars the project strives to contribute to the integration of two different schools of thought within phonology (a field of linguistics studying the sound systems of languages) when it comes to the storage of sounds in the mind. There are 'formal' approaches, typically occupied with abstractions, and 'usage-based' approaches, typically occupied with exemplars. The degree to which these two schools of thought communicate could be improved. Increased level of exchange of ideas and research results would facilitate the development of the field at large. Consequently, a unified approach, based on hybrid models of storage of sounds, is advocated.

The research to be carried out

Speech corpora of American English, Scottish English and Polish will be searched through with specialized open-source software to look for words whose acoustic properties can support either the need for abstractions or exemplars. The range of phenomena will include: the variation in the duration of fricative sounds (such as 's' in *speak* or 'th' in *thwart*) depending on the words in which these sounds occur, the reduction in the duration of words depending on whether a given word is typically predictable from its context, the weakening of the 'r' sound after vowels in Scottish English depending on an interaction between speakers knowing one another and what word they are using, and others. The properties of the audio signal will be measured automatically. Further information known from previous research to be relevant (such as how frequent a word is or how fast speakers are talking) will be added, and entered into statistical models. With the help of this way of treating this selection of phenomena, the role of abstractions and exemplars can be investigated.

For example, if the 'r' sound is weakened more when Scottish speakers talk to people they know than when talking to strangers, but if this is so only for some words and not for others, it will be evidence of storage of detailed information as exemplars (there is no switch from one speech style to another, but rather each word behaves differently). If, on the other hand, the 'r' sound is weakened when occurring after a set of vowels that can be said to form a class on account of having something in common with one another (e.g. they are all pronounced with the tongue raised), it will be evidence of storage of abstractions (we can talk about common features of sounds only after abstracting away from a great deal of detail). If both effects are discovered, hybrid models will be supported.

Reasons for choosing the research topic

The issue of storage of sounds in the mind as abstractions or exemplars is crucial for phonology. Too little integration of findings of research conducted from the two, often seen as opposing, schools of thought slows down progress in the field. This research project offers a contribution to an integration of the two by uncovering, with the same method applied to similar sets of data, evidence for the need for both abstractions and exemplars. Additionally, by showing how conversational speech data can be harnessed to weigh in on issues relevant to phonological theory, the project is hoped to contribute to further popularization of this innovative way of doing research.