LING 2220 LING 5220

## Phonetics II: Data Science

We've entered the age of Big Data in the study of spoken language, as in other areas of science and engineering. We have access to very large digital datasets in hundreds of languages, and it's easy to collect more. And modern hardware and software makes it increasingly easy to get empirical answers to theoretical, practical, and social questions.

In this course, you'll learn to conduct phonetics research using modern resources and methods, applying a Big Data perspective both to classic problems and to new ones. Topics include scripting and statistical techniques, automatic phonetic analysis, and integration of methods from speech technology.

The course will be organized around three parallel tracks: Readings, to introduce problems and types of solutions; Exercises, to teach relevant techniques; and Projects, either individual or group. The details of the readings and exercises will depend to some extent on the backgrounds and goals of the participants.

Term projects may be individual or group. These projects can be something a participant is already working on, or something new, and varied lists of possible projects will be available. Graduate students enrolled in 5220 will produce a description of their project suitable for submission to a conference; undergrads enrolled in 2220 may do so if they wish.

There will be two meetings per week: each meeting might be a discussion of readings, or an interactive session to work on an assigned exercise; or a lab session to help plan and carry out an individual or group project.

The course materials will all be on line – no textbook or other purchases are required.