Microphones and formant estimates

The study reported here arose from a request to evaluate different types of recording equipment for the LANCHART Project, a longitudinal study of language change with Danish as an example. One aim of the assignment was to ensure that the LANCHART corpus would be suitable for certain acoustic phonetic investigations.

In our survey we considered 3 different types of microphones: Head mounted miniature microphones, lapel-worn miniature microphones and full-size directional microphones placed in a microphone stand in front of the speaker.

The following four factors were considered in the evaluation of the suitability of the recordings provided by the microphones: 1) ease of transcription and 2) segmentation of the recordings as well as 3) estimation of fundamental frequency and 4) estimation of formants using LPC analysis.

It was fairly simple to judge by the first 3 criteria how each microphone performed, but with respect to the formant estimation it was not evident which microphone type provided the best recording.

One could simply assume that the microphone setup which provides the loudest speech signal relative to unwanted noise from the surroundings would provide for the best formant estimates. One could also assume that microphone setups which produce a lesser signal to noise ratio would provide for formant estimates of lower precision, but with comparable values when averaged over a large number of tokens.

However, informal experiments indicated that these assumptions did not hold true. Apparently LPC based formant estimates can be affected by a number of factors related to the recording circumstances which interact in complex ways, making for a source of error of unknown impact on formant estimates.

The question is: how big is the problem? Is it large enough to have practical consequences for the use of LPC based formant estimation as an analysis tool? As an attempt to answer this question, a more comprehensive experiment has been designed. It seems to us that what we need is some sort of neutral reference recording and knowledge about the consequences for formant estimation as we deviate from this ideal. Thus we compare formant estimates of recordings made in four locations with very different acoustic characteristics using four different microphones simultaneously. In total the recorded material covers: 4 microphones (2 head mounted, full-size directional microphone, and a dedicated measuring microphone), 4 locations (Anechoic chamber, recording studio, two private rooms), 2 male and 2 female adult speakers. The recorded material consists of read sentences and sustained vowels with f0-sweeps (for more details see Hansen & Pharao, 2006).

The results available so far indicate that also when averaging over a larger number of tokens, different microphone type and placement can lead to substantially different formant estimates. Some vowel qualities seem to be more easily affected, and the effect is more pronounced for female speakers. Differences in the range from 2 to 10% (and a few beyond) are all highly statistically significant.
References