Testing the Ecological Validity of Repetitive Speech

Greg Kochanski and Christina Orphanidou
Phonetics Laboratory, University of Oxford, UK

We gratefully acknowledge support from the UK's Economic and Social Research Council: RES-000-23-1094.

Repetitive speech has been used in experiments on speech rhythm and syllable structure; such research assumes that repetitive speech is reasonably "natural". In this study, we test that assumption. We ask subjects to read out phrases repetitively to the beat of a metronome and we compare their readings with individual productions of the same phrases from a randomized list. The comparison is based on a lexical discrimination distance, which is a data-driven spectral distortion measure that we have developed. It responds to linguistically relevant differences: pairs of speech sounds from equivalent places in the identical words have a small distance; unmatched pairs have a large distance.

Experimental Methods: Ten speakers produced a total of 2578 utterances from 828 different texts for the training data. The mean sentence length was 6.5 words. Examples: "He was not sure himself", "Talking of wandering."

For the science data, 9 subjects read out a set of 48 phrases, 4-6 syllables long. The sentences were first read-out from a randomized list, contiaining 5 copies of each of the 24 phrases and over 50% filler phrases. After this, the subjects were asked to read-out the sentences repetitively, to the beat of a metronome. Examples: "Under the desk." "Nothing matters."

Developing a measurement technique:

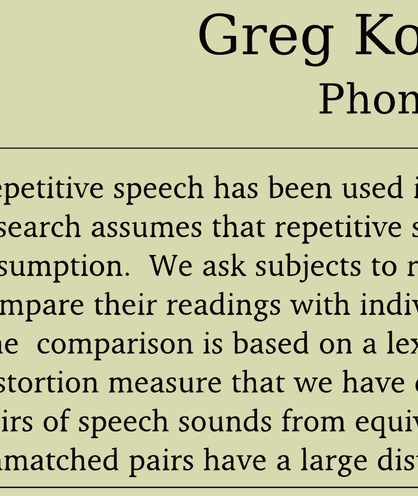
Choose a pair of utterances

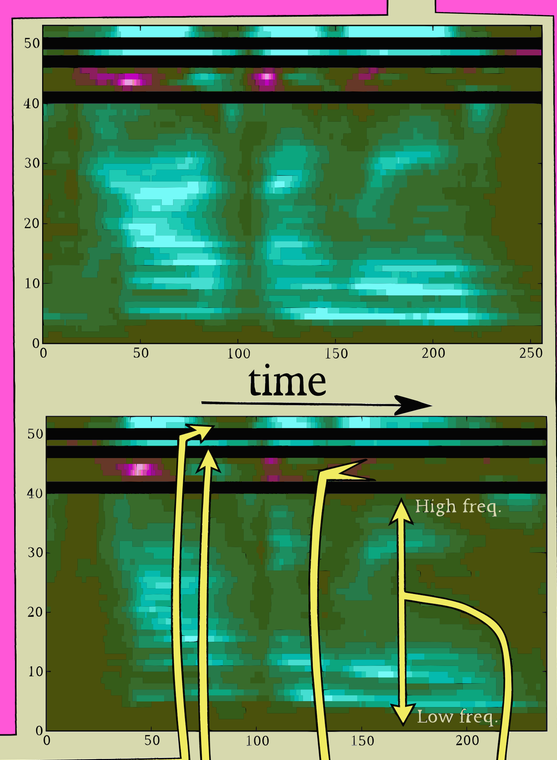
Compute audio description vector at each time

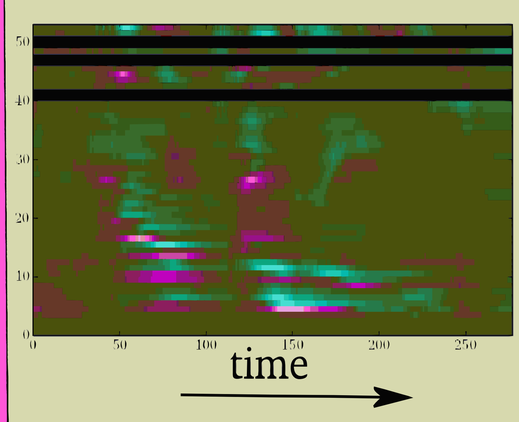
Match via dynamic time warp & subtract.

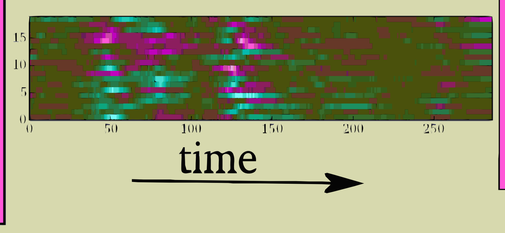
Projection (linear transform)

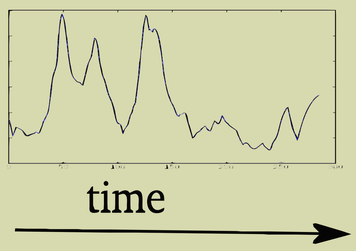
Take magnitude (to get distance)











These are two slightly different voicing detectors. They both work by computing an approximation to neural signals in the auditory nerve. A measure of the periodicity is computed in each band, and then it is summed across all frequencies.

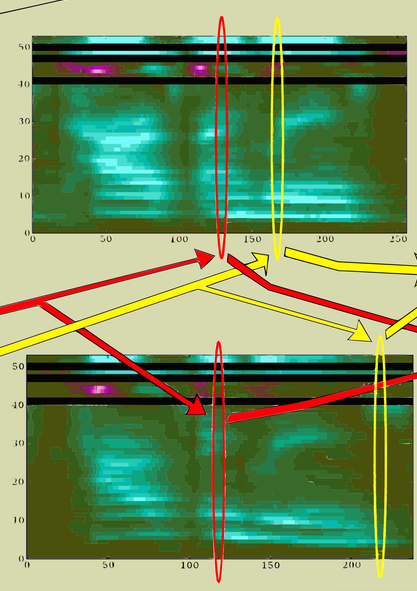
These are edge detectors with coarse frequency resolution: e.g. is the low-frequency specific loudness going up (magenta) or down (cyan)?

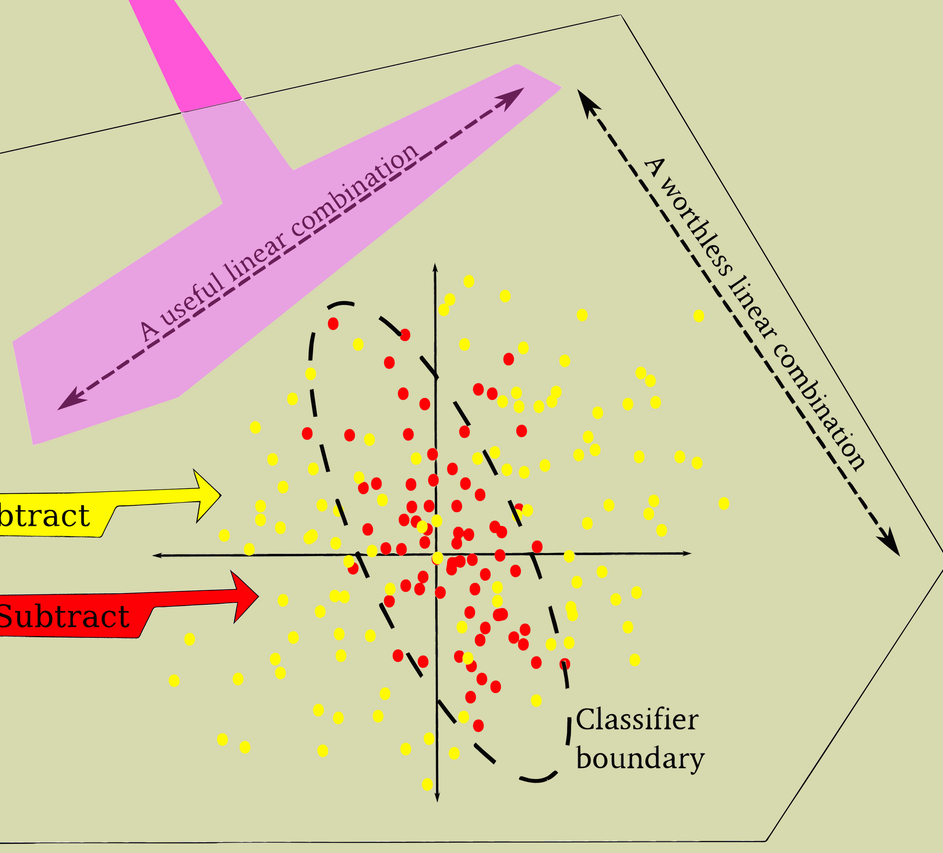
This is a perceptual power spectrum (specific loudness). The frequency resolution matches cochlear filtering. It has a cube-root intensity dependence, to match loudness perception. Then, it is divided by a mixture of the instantaneous loudness and the overall average loudness.

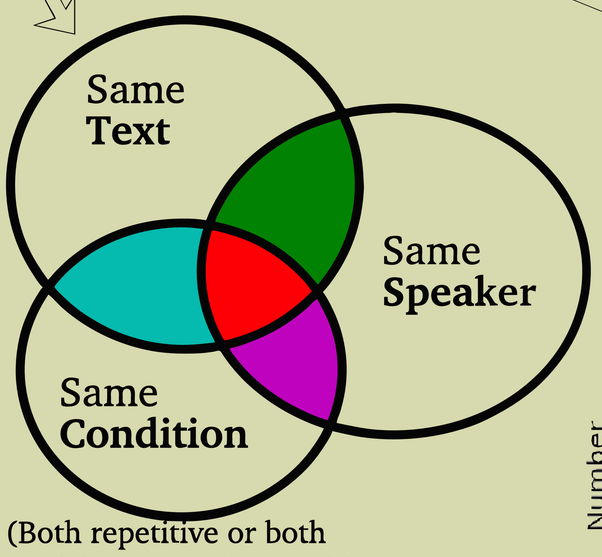
This is the entropy of the spectrum at each moment. Large (cyan) means a complex spectrum with peaks, small (green/brown) means a simple spectrum that is nearly flat.

For the experiment, we make four varieties of pairs.

For training, we always use pairs that come from the same texts. We make two classes:
MATCH: same point in the same word.
MISMATCH: at least 250ms away from a match.

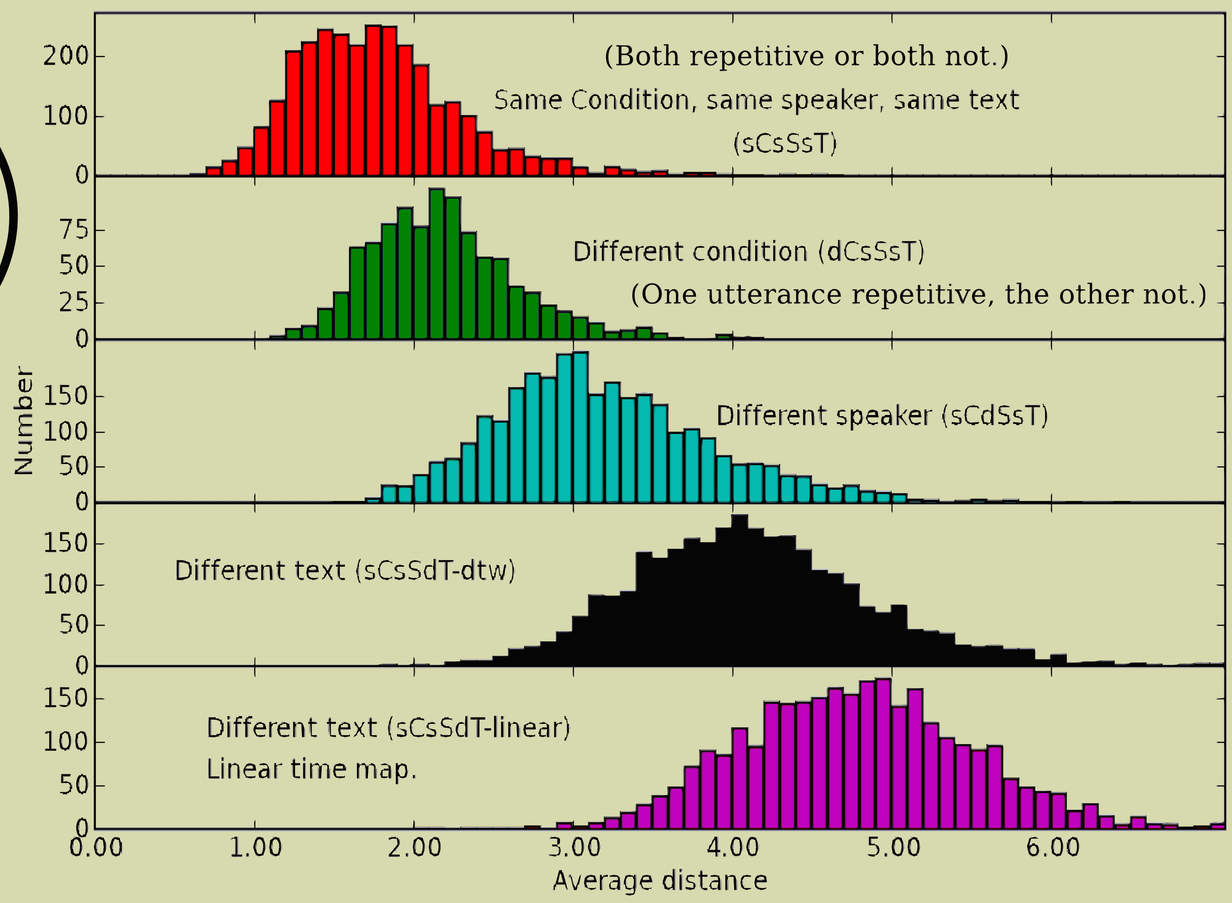






(Both repetitive or both read from a randomized list.

Measuring the average distance between repetitive and non-repetitive speech:



Conclusions:

1) We have developed a technique for measuring a linguistically-relevant distance between utterances. It separates lexically important differences from irrelevant inter-speaker differences and random pronunciation variation. Averaged over a short utterance, it can resolve pronunciation changes as small as 20% of a minimal pair distance.

2) We tested and calibrated it by comparing distances between pairs of utterances produced by (a) the same person, (b) different people, and (c) utterances produced from different texts: a<b<c, as expected.

3) We then used it to compare repetitive and non-repetitive utterances. The distance between repetitive and non-repetitive speech is smaller than utterance-to-utterance variability and substantially smaller than typical differences between speakers.

Thus, experiments can safely use repetitive speech*.

*(1) The results apply to the average spectral distance between utterances. Possibly, there are systematic differences in small areas that do not show up in the average.
(2) The experiment does not measure timing changes, so there could be differences in the duration of phones and/or syllables.

Download this poster from http://kochanski.org/gpk/papers/2007/icphs_poster.svg or .pdf
Download the full paper from http://kochanski.org/gpk/papers/2007/2007repeated.pdf