Autoregressive models

models

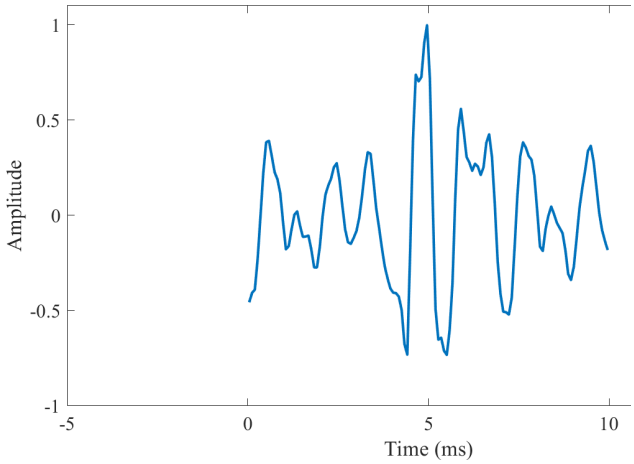
Look at the speech signal segment to the right. On a large scale it is hard to discern a structure, but on a small scale, the signal seems continuous. Speech signals typically have such structure that samples near in time to each other are similar in amplitude. Such structure is often called short-term temporal structure.

More specifically, samples of the signal are *correlated* with the preceding and following samples. Such structures are in statistics measured by covariance and correlation, defined for zero-mean variables x and y as

$$\text{covariance: } \sigma_{xy} = E[xy]$$
$$\text{correlation: } \rho_{xy} = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}},$$

where *E[ ]* is the expectation operator.

Short segment of speech



For a speech signal x_n, where *k* is the time-index, we would like to measure the correlation between two time-indices x_n and x_h. Since the structure which we are interested in appears when *n* and *h* are near each other, it is better to measure the correlation between x_n and x_{n-k}. The scalar *k* is known as the *lag*. Furthermore, we can assume that the correlation is uniform over all *n* within the segment. The self-correlation and -covariances, known as the *autocorrelation* and *autocovariance* are defined as

$$\text{autocovariance: } r_k = E_n[x_n x_{n-k}]$$
$$\text{autocorrelation: } c_k = \frac{E_n[x_n x_{n-k}]}{E_n[x_n^2]} = \frac{r_k}{r_0}.$$

The figure on the right illustrates the autocovariance of the above speech signal. We can immediately see that the short-time correlations are preserved - on a small scale, the autocovariance looks similar to the original speech signal. The oscillating structure is also accurately preserved.

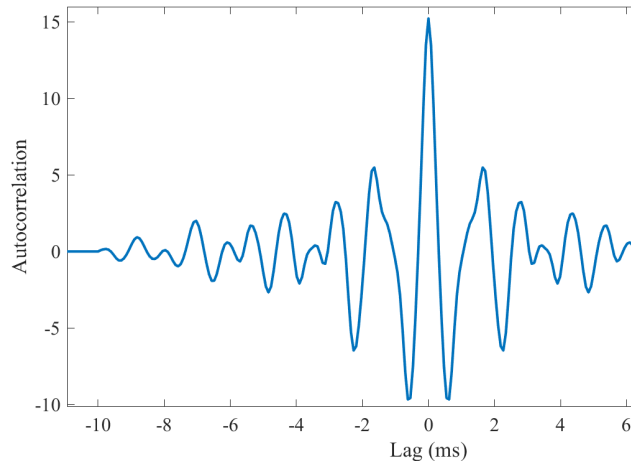
Because we assume that the signal is stationary, and as a consequence of the above formulations, we can readily see that autocovarinaces and -correlations are symmetric

$$r_k = E_n[x_n x_{n-k}] = E_n[x_{n+k} x_{n+k-k}] = E_n[x_{n+k} x_n] =$$

This symmetry is clearly visible in the figure to the right, where the curve is mirrored around lag 0.

The above formulas use the expectation operator *E[ ]* to define the autocovariance and -correlation. It is an abstract tool, which needs to be replaced by a proper estimator for practical implementations. Specifically, to estimate the autocovariance from a segment of length *N*, we use

The autocovariance of a speech segment



$$r_k \approx \frac{1}{N-1} \sum_{n=k+1}^{N-1} x_n x_{n-k}.$$

models

We can also make an on-line estimate of the autocovariance for sample position *n* with lag *k* as

$$\hat{r}_k(n) := \alpha x_n x_{n-k} + (1-\alpha)\hat{r}_k(n-1),$$

where α is a small positive constant which determines how rapidly the estimate converges.

It is often easier to work with vector notation instead of scalars, whereby we need the corresponding definitions for autocovariances. Suppose

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}.$$

We can then define the autocovariance matrix as

$$R_x := E[xx^T] = \begin{bmatrix} E[x_0^2] & E[x_0 x_1] & \dots & E[x \\ E[x_1 x_0] & E[x_1^2] & \dots & E[x \\ \vdots & \vdots & \ddots & \\ E[x_{N-1} x_0] & E[x_{N-1} x_1] & \dots & E \end{bmatrix}$$

Clearly R_X is thus a symmetric Toeplitz matrix. Moreover, since it is a product of *x* with itself, R_X is also positive (semi-)definite.

无标签