Mode locking of voiced speech reconstructed by mono-causally coupled phase dynamics

F.R. Drepper
Zentrallabor für Elektronik, Forschungszentrum Jülich GmbH

Due to the strong amplitude variation, characteristic for human speech, it is advantageous to detect the voicing of resonances in the articulatory cavity as a phase synchronization between the glottal dynamics and specific harmonics or formants of the sound pressure signal. The central analytic step of the "phaselet" reconstruction being introduced is the estimation of a set of coupled circle maps describing the time discrete phase dynamics of different frequency bands obtained from a filter bank with logarithmic frequency scale. The spontaneous frequency variation of the glottal sound source does no longer appear as a nasty disturbance of the periodicity and harmonicity of voiced speech but as an offer to detect the driver – response relationship between phase synchronized signals. In a companion paper (Drepper 2002) it is shown that voicing manifests itself as the existence of a single frequency band of the sound signal, whose phase is suited to replace the glottal phase in its role as the common driver.

Contrary to Ohm and Helmholtz the entrance step of human auditive cognition is assumed to be focussed on the dynamics of the phases of band limited oscillators generated by the filter bank of the basilar membrane. In particular it is assumed that acoustic objects are identified and discriminated according to specific features of the phase synchronization between different band specific oscillators. The features include the direction of the coupling leading to phase synchronization. Voiced objects are characterized by unidirectional coupling of a single driver to a whole set of response oscillators.

To be able to discriminate the more complex forms of phase synchronization realized in voiced human speech (companion paper) and in particular to detect its characteristic unidirectional coupling, a phenomenological transfer function model is needed, which is focussed on the reconstruction of coordinated phase dynamics.

Phaselet reconstruction of a broadband response driven by a bandlimited oscillator

To separate different frequency bands of the response signal a logarithmic scale filter bank is used, which bears some resemblance to wavelet decomposition - a feature which can be exploited to obtain efficient code for the filter bank. The concern about redundancy is central to both approaches however they differ in their strategy of abstraction. Whereas wavelet decomposition adjusts the sampling in time and frequency (scale) such that redundancy is avoided, the present approach uses a deliberate over-sampling, since the redundancy is used to estimate coupled deterministic dynamics of the band specific phases. This applies to redundancy in form of autocorrelation within each band as well as to cross correlation between different bands. In fact the optimal mesh size of the grid used in a phaselet reconstruction turns out to be twice as dense as the one of the wavelet decomposition. This applies to the grid in time as well as to the frequency grid.

Due to the a-causal, infinite impulse response property of the Hilbert transform, Hilbert phases are not suited for a deterministic reconstruction. For band limited signals the Hilbert phase can be substituted by a (topologically equivalent) finite impulse response alterna tive.
A phaselet $\{\varphi_{j,n} \mid n=1,...,N_j\}$ is defined as a FIR phase of the output $\{X_{j,m} \mid m=1,...,M\}$ of the filterbank, in particular

$$\varphi_{j,n} = \arctan(X_{j,(n+1)\Delta}, X_{j,n\Delta}) \quad \text{for } (n=1,...,N_j), \qquad (1a)$$

where Δ represents the band specific sample rate chosen as $\Delta = T/4$, where T represents the average period length of band *j*. The bivariate function $\arctan(Y,X)$ extends the range of mono-variate function $\arctan(Y/X)$ to the full circle from $-\pi$ to π and $N_j = [M/\Delta]$ denotes the band specific number of phases. The related amplitudes or radii $r_{j,n}$ are defined as,

$$r_{j,n} = \sqrt{X^2_{j,(n+1)\Delta} + X^2_{j,n\Delta}}\ . \qquad (1b)$$

In the limiting case of no coupling to other modes, each band limited signal is chosen as a stable linear second order autoregressive process. In the limiting case of a strong coupling to a driving phase, the dynamics of a phaselet is designed to converge to a unique synchronization manifold defined by a unique, continuous relation to the simultaneous driving phase. This leads to the following linear amplitude, nonlinear phase transfer function model for each band limited signal with index *j*,

$$X_{j,(n+2)\Delta} = -a_j X_{j,(n+1)\Delta} - b_j X_{j,n\Delta} + r_{j,n} G_j(\psi_{n\Delta}) + \sigma_j \xi_{j,n+1} \qquad (2a)$$

where $\sigma_j \xi_{j,n+1}$ represents the momentary innovation, which for simplicity is assumed to represent an independently and identically distributed Gaussian random variable with standard deviation σ_j. $G_j(\psi_{n\Delta})$ can be interpreted as the driver phase specific directed cross

impact strength. Since every function of phases must be periodic with period 2π, $G_j(\psi_{n\Delta})$ can be approximated by a finite Fourier series,

$$G_j(\psi_{n\Delta}) = \sum_{k=0}^{K_j} \left(c_{j,k} \cos(k\psi_{n\Delta}) + d_{j,k} \sin(k\psi_{n\Delta}) \right) \tag{2b}$$

The phenomenological parameters a_j, $c_{j,k}$ and $d_{j,k}$ are estimated by standard multiple linear regression. The two dimensional autoregressive state space is transformed to polar coordinates as defined in (1a) and (1b)

$$\tan\varphi_{j,n+1} \sin\varphi_{j,n} = -a_j \sin\varphi_{j,n} - b_j \cos\varphi_{j,n} + G_j(\psi_{n\Delta}) + \frac{\sigma_j}{r_{j,n}} \xi_{j,n+1}$$

$$r_{j,n+1} = \frac{\sin(\varphi_{j,n})}{\cos(\varphi_{j,n+1})} r_{j,n}. \tag{4a}$$

When implicit equation (3) is solved for $\varphi_{j,n+1}$, the explicit form of the time discrete stochastic phaselet process for band *j* is obtained,

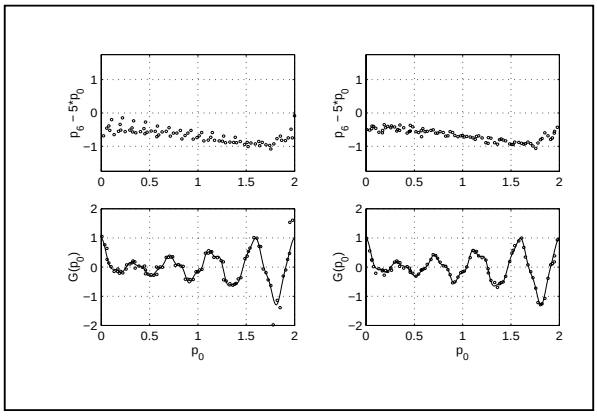
$$\varphi_{j,n+1} = \arctan\left(\begin{array}{l} G_j(\psi_{n\Delta}) - a_j \sin\varphi_{j,n} - b_j \cos\varphi_{j,n} \\ + \frac{\sigma_j}{r_{j,n}} \xi_{j,n+1}, \quad \sin\varphi_{j,n} \end{array} \right) \tag{4b}$$

.

Note that due to the direct phase to phase coupling the deterministic part of (4b) is independent of the complementary amplitude process defined in (4a). However the corresponding stochastic process is coupled to the complementary amplitude process, because the innovation evidences a $1/r_{j,n}$ dependent standard deviation (ARCH model), a feature which turns out to be important for the reconstruction of beat phenomena, characteristic for higher bands. Reconstructed amplitude and phase pairs can be converted to reconstructed band limited signals $X_{j,n\Delta}$. Finally the lowly redundant band specific signals can be up sampled to be represented on the common elementary time grid (by making use of their band limitation), so that they can be summed up to obtain a reconstruction of the original sound.

Results

The present study uses a simultaneous recording of the electro-glottogram and the sound pressure signal of voiced human speech (English wovel ae). It does not come as a big surprise that most frequency bands of the filter bank decomposition of the sound signal (up to about 2.5 KHz) evidence a near perfect phase synchronization or phase locking with respect to the phase of the glottal dynamics. The upper half of Fig. 1 shows the relative phase of a band with a (1:5) phase locking. In this representation the so called identical phase synchronization appears as a horizontal line. The lower half shows a statistics which is obtained, when the deterministic part of equation (2) is solved for $G_j(\psi_{n\Delta})$. In combination with the estimated function $G_j(\psi_{n\Delta})$ this graph is suited to disclose the driver phase specific ratio of the unexplained to the explained variance of model (4b). The left hand side of Fig. 1 has been generated with the input to the regression (corresponding to 30 ms recording), whereas the right hand side gives the same graphs for the reconstruction. The successful reconstruction of a (continuous and invertible) synchronization manifold can be seen as (double) evidence of phase synchronisation. The absence of phase synchronisation would result in absence of the dimension reduction in the subspace of the phases, which would lead to an insignificant estimation of the model parameters and to an attractor at a spurious position, which is not correlated with the original data.



Relative phase and driver phase specific cross impact strength $G_j(\psi\ )$, ($p_j \equiv \varphi_{j,n}$ and $p_0 \equiv \psi$)

The detection of phase synchronization can also be achieved formally by checking the negativity of the conditional Liapunov exponent in phase direction (Pecora and Carroll 1990, companion paper) – conditioned on the time evolution of the driving phase. For sustained voiced sounds both exponents are significantly negative. For voiced sounds with instationary amplitude the Liapunov exponent in the amplitude direction either approaches zero or becomes positive. In this case the dimension of the synchronization manifold increases by one. In fact it is the direct phase to phase coupling of the phaselets, which leads to high significance scores for the estimation of their parameters even in the instationary case and which makes the phaselets well suited to analyse human speech, characterized by simultaneous occurence of low amplitude coordinated jitter and highly instationary amplitudes (jimmer).

I extend my thanks for helpful discussions to
N. Stollenwerk, London, G. Reinerts, Oxford,
P. Grassberger and H. Halling, Jülich.

Drepper F.R., *Fortschritte der Akustik – DAGA'02*, (2002)
Drepper F.R., *Phys. Rev. E* **62**, 6376-6382, (2000)
Pecora L.M. and T.L. Caroll, *Phys. Rev. Lett.* **64**, 821 (1990)
Rulkov N.F.et al., *Phys. Rev. E* **51**, 980-994 (1995)
Email: f.drepper@fz-juelich.de