Unidirectionality of voicing detected by phaselet reconstruction

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

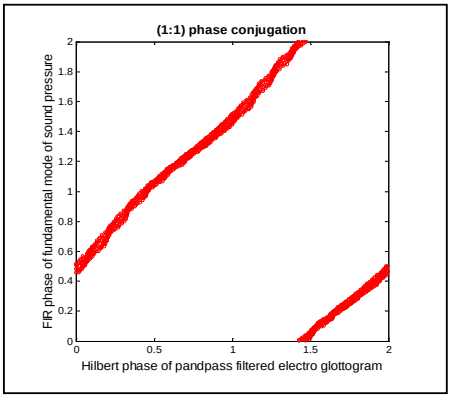
Voiced human speech can be interpreted as an acoustically transmitted sign language, the symbols of which are defined i.p. by size and shape parameters of the articulatory cavity on the transmitting side and by related mode locking phenomena of the resonant sound pressure signal on the receiving side. In a companion paper (Drepper 2002) it is shown that the resonant modes evidence a phase synchronization relationship to the phase of the glottal dynamics and that the obvious driver - response relationship can be identified by estimating a set of driven circle maps describing the time discrete response dynamics of the phases of different frequency bands obtained from a filter bank with logarithmic frequency scale. The present study demonstrates 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. The reconstruction of the latter driver response dynamics may be used for perceptionally equivalent sound reconstruction and as an acoustic object identification and discrimination step in the entrance level of automatic speech recognition.

The old concept (Huygens 1673) of synchronization has recently received two important extensions opening the notion to non-identical and aperiodic, coupled oscillators: generalized synchronization of unidirectionally coupled oscillators (Rulkov 1995) and phase synchronization (Rosenblum 1996). The central notion of generalized synchronization in driver - response systems is a unique and continuous map which relates a state of the response system to the simultaneous state of the driver. The continuous map defining the synchronization manifold is required to establish a unique relation for a complete set of state variables of the response oscillator. The central notion of phase synchronization or phase locking is the two sided boundedness of the "relative phase", which is defined as the difference between specific integer multiples of two unwrapped phases (Rosenblum 1996). In the case of phase synchronisation in a driver - response system the two concepts can be united by interpreting phase synchronisation as a generalized synchronization, where the range of the mentioned unique map is limited to one dimension - the phase of the response oscillator (Drepper 2000). However the continuous map relating the simultaneous states has to fulfil an additional invertibility property. In the case of a (1:1) phase synchronization, the response phase has to be related to the phase of the driver by a homeomorphic (invertible and continuous) map, which establishes an equivalence relation (conjugation) between the two phases. In the more general case of a (1:n) phase locking the response phase becomes a conjugate of the nth multiple of the driver phase.

The application of the phaselet reconstruction being presented is based on the reciprocity, transitivity and comparative time invariance of several phase conjugations related to resonances in the articulatory cavity. Examples in place are the (1:n) phase conjugation between the glottal dynamics as observed in the electro-glottogram and related resonant sound pressure modes within the articulatory cavity or the (1:1) conjugations between the latter modes and corresponding ones obtained from a filter bank decomposition of the sound pressure signal reaching the receiver. Such conjugations may be difficult to prove a priori. However a successful phaselet reconstruction can be seen as a selfconsistent evidence for the existence of an uninterrupted chain of phase conjugations (companion paper).

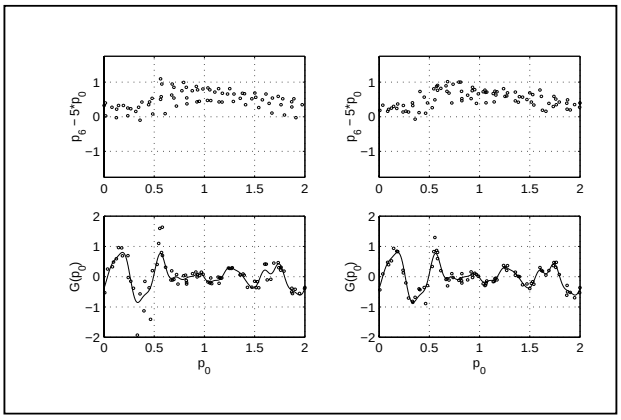
It is a remarkable feature of conjugate dynamical systems that there exist topological invariants (Kantz 1997) like conditional Liapunov exponents and the periodicity and/or dimension of attractors on the synchronization manifold (Appendix A). Since they are guaranteed to have the same value on the transmitter and on the receiver side, these invariants are particularly suited to carry relevant information. Indeed it is remarkable to which extend human speech is characterized by a pronounced variation of the synchronization related topological invariants. Voicing is characterized by a negative conditional Liapunov exponent in the phase direction. The coexistence of specific periodicities of different phaselets is known to be characteristic for vowels. The plosives are obviously characterized by a positive conditional Liapunov exponent in the amplitude direction and for stop consonants this exponent can be expected to be strongly negative.

Whereas a phase conjugation alone contains no hint on the direction of the coupling, the phaselet reconstruction is suited to disclose this information. Phaselet reconstructions make intelligent use of the reduced dimension of the attractors on the characteristic synchronization manifold, defined by the phase conjugation. The nonlinearity of the dynamics of the phaselets, has been deliberately restricted to a phase to phase coupling term, which is linear in the phenomenological parameters. This tamed nonlinearity supports a safe and easy estimation of the model parameters. In spite of this restriction, the potential complexity of phaselet attractors exceeds the one of the linear source and filter model by far.

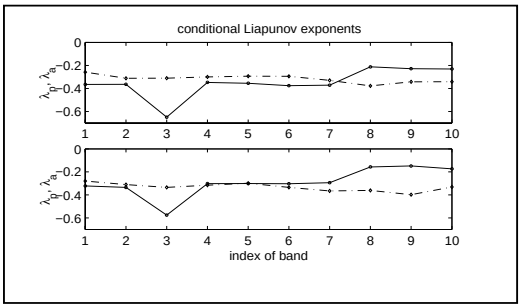


As is shown in Fig. 1 the phase of the glottal dynamics as recorded in the electro-glottogram and the phase of the fundamental mode of the sound pressure signal are related by a perfect conjugation (phase synchronization). When combined with the findings of Fig.1 of the companion paper, the topological equivalence of these two phases suggests the detection of the directed causality and phase synchronization of the oscillatory modes within the articulatory cavity as an auto driving and auto phase synchronization between different frequency bands of a single sound signal reaching the receiver.

Fig. 2 demonstrates the auto phase synchronization for the empirical data (left side) and for the phaselet reconstruction based on the driving by the fundamental phase of the sound signal (right side).



When comparing Fig. 2 to Fig. 1 of the companion paper a good qualitative agreement of the synchronization manifolds and the driver phase specific cross impact functions can be observed. Note that the near constant phase shift between the two drivers - to be seen in Fig. 1 of the present paper - is irrelevant for the success of the proposed substitution. Fig. 3 demonstrates to which extend the two conditional Liapunov exponents agree for the two types of driving. In the case of the conditional Liapunov exponent in the phase direction the range of good agreement is restricted to frequencies up to about 1.5 KHz (band 8) .



The use of the linear source and filter model in the CELP codec has no severe disadvantage for speech reconstruction, since its inability to describe the more complex attractors can be easily compensated by frequent updates. However in the case of speech recognition an important possibility of noise rejection or suppression gets lost, when the detection and distinction of the more complex variants of phase synchronization is shifted from the entrance level to the next higher level, e.g. the HMM level. Thus the phaselet reconstruction based on the driving by the fundamental mode may turn out as a useful new first step to identify and discriminate speech and music in noisy environments.

The hypothesis that the basilar membrane performs an acoustic object identification and classification by some kind of phaselet reconstruction would open up a new perspective to understand the difference in qualitative detail between the subjective richness of perceived human speech on the one hand side and the objective vagueness of the visualized spectrogram on the other side. The latter instrument of analysis is known to be based on a linear system description. The phaselet reconstruction method being presented can offer help to perform the necessary psychoacoustic experiments towards falsification or further qualification of the hypothesized human auditive abilities and to supplement or replace the feature vector used in automatic speech recognition, as necessary.

Appendix A

Liapunov exponents describe the long time dynamics of the difference between two trajectories, defined by an arbitrarily small and instant perturbation of a reference trajectory. Since the state space of a phaselet oscillator is two-dimensional, in particular φ_n , r_n , this difference will have two components as well. The assumption of an infinitesimally small perturbation allows the linearization of the dynamics around the reference trajectory, leading to the Jakobi matrix, which can easily be derived from equation (4a) and (4b) of the companion paper. The time evolution of the perturbation is obtained as the repeated product of the Jakobi matrix, which fortunately turns out to be a lower triangular matrix with the useful property that the diagonal of their product results in the products of the diagonal elements of the single Jakobi matrices. The two Liapunov exponents describing the average logarithm of the dilation or shrinking factor in the direction of the phases, λ_φ , and amplitudes, λ_a , can thus be obtained as

$$\lambda_\varphi = \frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{\left| b - G(\psi_{n\Delta})\cos\varphi_n \right|}{\sin^2\varphi_n + \left(G(\psi_{n\Delta}) - a\sin\varphi_n - b\cos\varphi_n\right)^2} \right)$$

$$\lambda_a = \frac{1}{N}\sum_{n=1}^{N} \ln\left(\left| \frac{\sin(\varphi_n)}{\cos(\varphi_{n+1})} \right| \right).$$

For stationary voiced sound both conditional Liapunov exponents are negative. In this case the initial perturbation converges to zero. That means that the long time dynamics of a phaselet becomes independent of its initial condition as long as this initial condition belongs to the corresponding basin of attraction. However its dynamics remains dependent on the dynamics of the driver. In general this dependence can include present and past states of the driver. In the case of a 1:n synchronization or phase locking this dependence degenerates to the present state of the driver. Note that the case of an m:n synchronization with $m \neq 1$ corresponds to an active role of past states of the driver.

For voiced sounds with instationary amplitude the conditional Liapunov exponent λ_a deviates strongly from the one in phase direction. For a diffusion type instationarity the second Liapunov exponent approaches zero and for exponentially increasing amplitudes (plosives) it becomes positive. In both cases the dimension of the synchronization manifold increases by one.

I extend my thanks for helpful discussions to V. Hohmann, Oldenburg and C. Hoelper, Aachen.

Hugenii C., Horoloquium Oscillatorium. *Parisis*, France (1673)
Rulkov N.F.et al., *Phys. Rev. E* **51**, 980-994 (1995)
Rosenblum M.G., et al., *Phys. Rev. Lett.* **76**, 1804 (1996)
Drepper F.R., *Phys. Rev. E* **62**, 6376-6382, (2000)
Drepper F.R., *Fortschritte der Akustik – DAGA'02*, (2002)
Kantz H., T. Schreiber, Nonlinear time series analysis, Cambridge Univ. Press (1997)