SPEECH PRODUCTION AND CHAOS

Hans–Peter Bernhard and Gernot Kubin

Institut für Nachrichtentechnik und Hochfrequenztechnik,
Technische Universität Wien,
Gusshausstrasse 25/389, A–1040 Vienna, Austria

ABSTRACT

The apparent irregularity of speech signals is generally considered the effect of pure randomness or of intended time-varying control. This paper presents a new explanation in terms of chaos which origins from deterministic, time-invariant, nonlinear systems. Experiments with sustained German vowels support this model for the temporal fine structure of speech.

1 INTRODUCTION

The present study is devoted to an *experimental investigation* whether the chaotic paradigm is useful for speech production modeling or not. Strange enough, only little research is reported in this area [3,4,6] although the variability of the speech signal suggests itself to an explanation in terms of chaos. This is further motivated from three observations in speech acoustics:

- If speech sounds are produced *unvoiced* the primary source consists in turbulent air flow originating at constrictions of the vocal tract [1, pp. 53–58]. Turbulent flow is one of the most extensively studied patterns of chaotic behaviour in fluid mechanics.
- If speech sounds are produced *voiced* there exists a nonlinear coupling between the sound source and the vocal tract [1, pp. 41–53, pp. 246–259].
- There is some experimental evidence that the standard linearized acoustic theory of speech production may not be fully adequate as the actual air stream mechanisms are much more complicated than generally believed [7].

2 ANALYSIS OF CHAOTIC SIGNALS

While conventional analysis resorts to the assumption of an underlying stochastic process to explain the irregular behaviour of observed signals, chaos is an inherent property of purely deterministic, nonlinear systems. From that, analysis of *chaotic signals* requires to a certain extent knowledge of the *system* producing the observations.

Reconstruction of Attractors from a Signal. Following a theorem by Takens [5, p. 191], it is possible to reconstruct the attractors of the underlying system from a simple procedure using a single output signal $s(t)$. To this end, equidistant samples $s(t + kT), k = 0, \ldots, N - 1$ are collected into an N-dimensional vector and the trajectory of this vector is plotted in N-dimensional state space.

Measurement of Fractal Dimension. There are various definitions for

the *fractal* dimension of a geometrical object [5, pp. 167–191]. We have implemented an estimator for the *correlation dimension* of the reconstructed attractor.

Measurement of Information Production Rate. The unpredictable behaviour of chaotic systems corresponds to a steady production of new information which is related to its *Ljapunov exponents* [5, pp. 73–81]. We follow an independent approach developed by Fraser [2] which computes the *mutual information* between a signal sample and a vector of delayed samples. This quantity defines the *marginal redundancy* $R(T, N)$ of the sample, i.e. how many bits of the sample are predictable from the past N samples with delay T. It exhibits three distinctive properties for chaotic signals:

- An increase of N makes the redundancy $R(T, N)$ reach a saturation line.
- This saturation line decreases with increasinf delay T. Its slope is the information production rate of the underlying system in bits/sec.
- The delay T can be chosen such that the information conveyed by the reconstructed attractor is maximized.

3 EXPERIMENTS WITH SPEECH SIGNALS

Our experiments cover long vowels [a:, e:, i:, o:, u:] spoken by three male native speakers of German, of age 24 to 30. Although the vowels were embedded in carrier sentences, they were sustained over several seconds to provide sufficiently long stationary data for further analysis. The recordings were made in an anechoic chamber and digitized with 16 bit resolution and 32 kHz sampling rate. The analysis software is implemented on an IBM PC-AT 286 for testing and visualization purposes while the more extensive batch-mode analysis runs were done on a MicroVax II minicomputer.

Results. Figure 1 shows a two-dimensional projection of the attractor corresponding to the vowel [i:] by speaker 'F'. The optimal delay time $T = 0.94$ msec. The graph displays three periodicities: (1) the trajectory almost replicates itself after every pitch period; (2) the first formant frequency is roughly twice the pitch frequency, so the overall structure of the graph is similar to a folded 8 (two cycles per 'pitch orbit'); (3) the second formant frequency is about 23 times the pitch frequency, so their are 23 small loops superimposed on the pitch orbit. From the reconstructed attractor, the correlation dimension is estimated as 1.7. This fractional value reflects the observation that the signal is *almost* periodic (otherwise it would be 1).

The overall structure of the attractor is maintained for other speakers. Figure 2 displays the attractor of vowel [i:] by speaker 'B'. It is characterized by the same looping *structure* while the *amplitude* of the second-formant loops is significiantly reduced due to power differences in the respective formants of the two speakers.

Figure 3 explains the importance of the right choice for the delay time T. While Figure 1 is obtained with the optimal T-value, this figure shows again vowel [i:] by speaker 'F' but with $T = 1.88$ msec. The optimal T-value produces results which are easier to read and, most interestingly, this value is rather independent of the individual speaker or vowel under consideration. *The optimal delay time is on the order of 1 msec.*

Redundancy analysis of the attractor of Figure 1 delivers Figure 4. On display are the marginal redundancies $R(T,N)$ for $N = 1,2,3$ and T between 0 and 500 sampling intervals (i.e. 0 to 15.6 msec). The curves saturate already for $N = 2$ so that $N = 3$ does not increase the predictable amount of information significantly. Prior to saturation (for $N = 1$), unmodeled periodicities are still visible as local maxima of the curve, note e.g. the pronounced peak due to the pitch at $T = 7.9$ msec. The slope of the saturation line is 0.94 bit over one pitch period. This is the information production rate of the underlying system.

Figures 5 and 6 show as a further example the reconstructed attractor and redundancy plots for the vowel [a:] by speaker 'G'. While differing slightly in the numerical values, the general picture found in the previous figures is corroborated. For comparison, the pertaining measurement values are listed here: correlation dimension 1.5, optimal delay time $T = 0.94$ msec, information production rate 0.9 bit over one pitch period.

4 CONCLUSION

Albeit the experimental basis is still rather limited, preliminary conclusions may be drawn:

- The investigated vowels can be interpreted in terms of deterministic chaos. This is supported through the positive information production rate of roughly one bit per pitch period and the fractional value of the correlation dimension between 1 and 2.
- As only sustained vowels with constant pitch have been investigated, the observed chaotic behaviour can only be related to the temporal fine structure of speech that changes from pitch period to period. No claim about the dynamic behaviour of spectral or articulatory parameters is made.
- The gross shape of the reconstructed attractors can be interpreted in terms of standard phonetic theory. Their 'strangeness', however, goes beyond such theory which offers only explanations in terms of randomness or time-varying control.

REFERENCES

[1] J.L. FLANAGAN. *Speech Analysis Synthesis and Perception.* Springer–Verlag, Berlin, 1972(2).

[2] A.M. FRASER. "Information and entropy in strange attractors," *IEEE Trans. on Inf. Theory*, IT-35(2):245–262, 1989.

[3] A. KUMAR and S.K. MULLICK. "Speech signal modeling *à la* chaos," *IEEE 1990 Dig. Sig. Proc. Workshop*, paper 1.8, New Paltz (NY), Sept. 1990.

[4] P. MARAGOS and K.L. YOUNG. "Fractal excitation signals for CELP speech coders," *Proc. ICASSP'90*, 669–672, Albuquerque (NM), Apr. 1990.

[5] T.S. PARKER and L.O. CHUA. *Practical Numerical Algorithms for Chaotic Systems.* Springer–Verlag, Berlin, 1989.

[6] T.F. QUATIERI and E.M. HOFSTETTER. "Short-time signal representation by nonlinear difference equations," *Proc. ICASSP'90*, 1551–1554, Albuquerque (NM), Apr. 1990.

[7] H.M. TEAGER and S.M. TEAGER. "Evidence for nonlinear production mechanisms in the vocal tract," in *Speech Production and Speech Modelling*, Kluwer, Dordrecht, 1990.

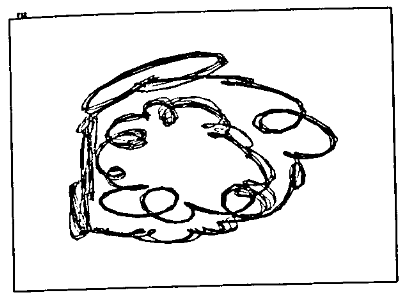

Figure 1: Attractor of vowel [i:] by speaker 'F' ($T = 0.94$ msec).

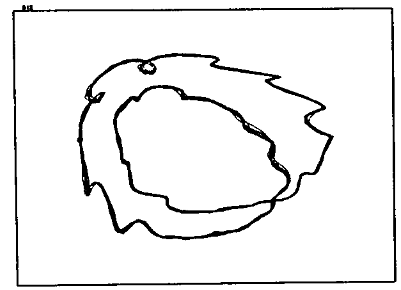

Figure 2: Attractor of vowel [i:] by speaker 'B' ($T = 0.94$ msec).

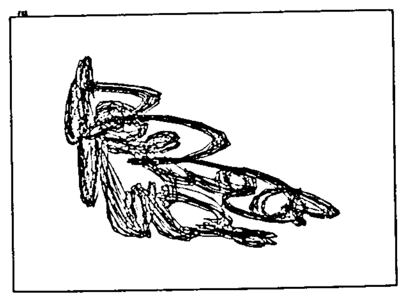

Figure 3: Attractor of vowel [i:] by speaker 'F'($T = 1.88$ msec).

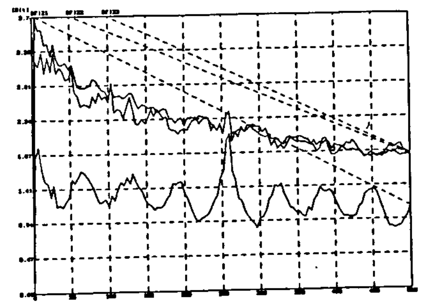

Figure 4: Marginal redundancy $R(T,N)$ of vowel [i:] by speaker 'F'.

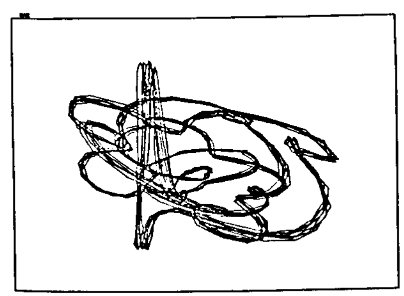

Figure 5: Attractor of vowel [a:] by speaker 'G' ($T = 0.94$ msec).

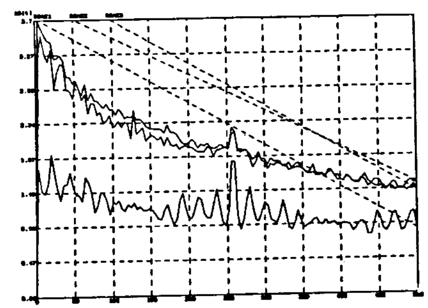

Figure 6: Marginal redundancy $R(T,N)$ of vowel [a:] by speaker 'G'.