hase representations of acoustic speech waveforms

Eric Keller

1 INTRODUCTION

Phase representations are commonly used in dynamic systems (Abraham & Shaw [1]). They offer summary glimpses at the complete, repeating behaviour of a given system. Such representations have proven their usefulness in speech kinematics, particularly with respect to their ability to capture phase relationships between interacting (co-articulating) muscular structures (e.g. Kelso & Tuller [4]). In this paper we briefly examine phase representations of acoustic speech signals.

2 METHOD

There are two ways to obtain a phase representation:

i For each sample in the signal, calculate the amplitude difference between the current and the previous sample, and plot this type of velocity against the measured amplitude for the sample. Thus

$$x_k = x(k) - x(k-1), \qquad y_k = x(k) \tag{1}$$

where x_k is the velocity at the current sample derived from the difference between samples $x(k)$ and $x(k\text{-}1)$, and y_k is the amplitude of the current sample obtained from sample $x(k)$.

ii With reference to samples in the signal, calculate an approximation to instantaneous velocity first (e.g. a spline, de Boor [2]) and plot this type of velocity against measured amplitude. Thus

$$x_k = s'(t_k), \qquad y_k = x(k) \tag{2}$$

where x_k in the phase plot represents one point in the first derivative $s'(t_k)$ of a sampled cubic function that approximates the signal's sample path $(s\,(t_k))$; y_k is the amplitude of the current sample obtained from sample $x(k)$.

The first approach is useful for a rough-and-ready first glance at the phase characteristics of a given signal segment. The second approach is to be preferred for the suppression of quantisation error or for filtering out high-frequency components.

Since phase plots involve velocity as their first parameter, they are different from somewhat similar-looking two-dimensional signal representations popularised some 30 years ago by Lerner [5]. The latter involve as their first parameter either the Fourier or the Hilbert transform (Demars [3]). Despite this difference in derivation, phase plots share some of the interesting properties of Lerner's polar plots, such as their relative lack of sensitivity to temporal parameters like sampling frequency or time scale. A further similarity is that phase plots retain time information without being particularly sensitive to it.

Figure 1 shows the derivation of a phase representation (or "phase portrait") from a time-amplitude plot for a sinusoidal cycle. As the signal rises, velocity first increases and then de-

---

Visual Representations of Speech Signals: Martin Cooke, Steve Beet and Malcolm Crawford (eds.)

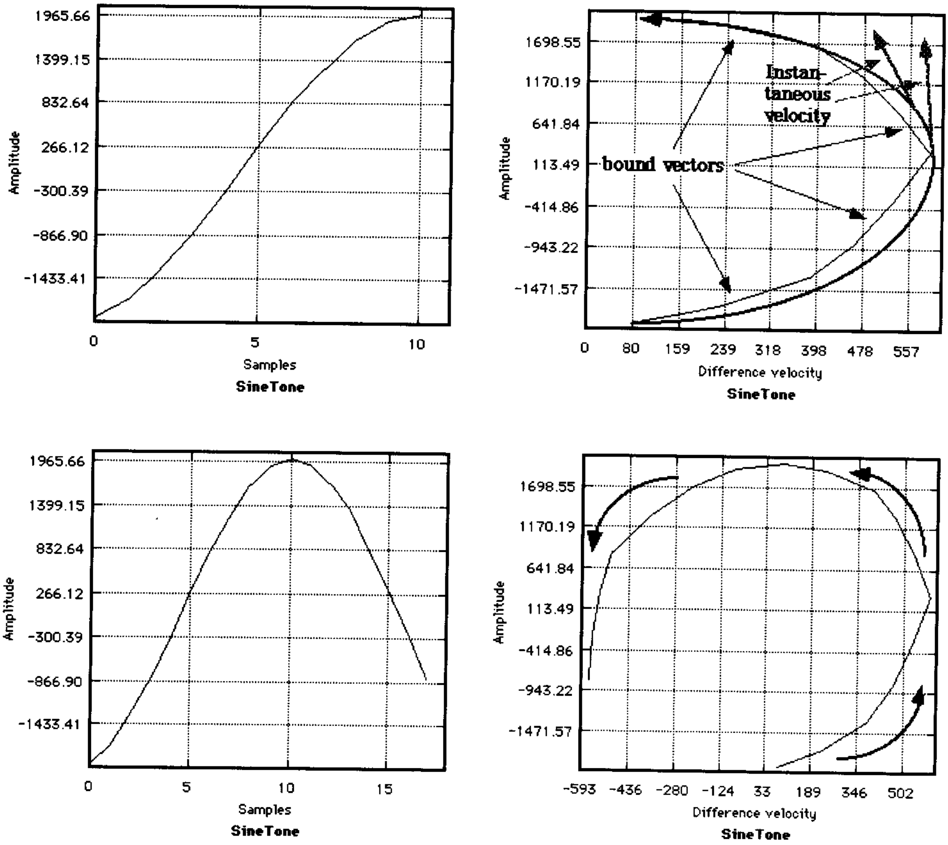

Fig. 1 Time-amplitude and phase representations of segments of a sinusoidal cycle. In the upper right figure, the rising portion of the sine wave cycle is translated into a phase portrait by the plotting of bound vectors, or of differences between adjoining samples (thin line). The assumed underlying smooth phase movement is shown as a thicker line.

creases. In the negative-moving signal, velocity first moves from zero into the negative domain and then back to zero. A circle is thus derived. In dynamic systems, velocity is often shown on the y-axis. However to preserve the similarity to the time-amplitude signal, we have chosen to show velocity on the x-axis and to keep amplitude on the y-axis.

In dynamic system terms, the time-amplitude plot shows a *state space* (i.e. a series of successive states of the system), while a phase plot shows a *vector field* (i.e. an area made up of vectors, or *directions* that the system moves into). In fig. 2, the relationship between time-amplitude and phase representations is illustrated with respect to a small segment from the /a/ sound of “can’t” in ‘clean.dip’.

The superimposition of vectors illustrates well the fundamental tendencies of the sound pressure system during the production of a vowel. Of particular interest in dynamic system theory are centres that appear to attract and repel the signal repeatedly (so-called attractor and repellent basins). At these points, the system shows a loss and regaining of energy. A basin may act like an attractor at the outer reaches of the system and like a repellent close to its basin. The line between the two areas is known as the *limit cycle*.

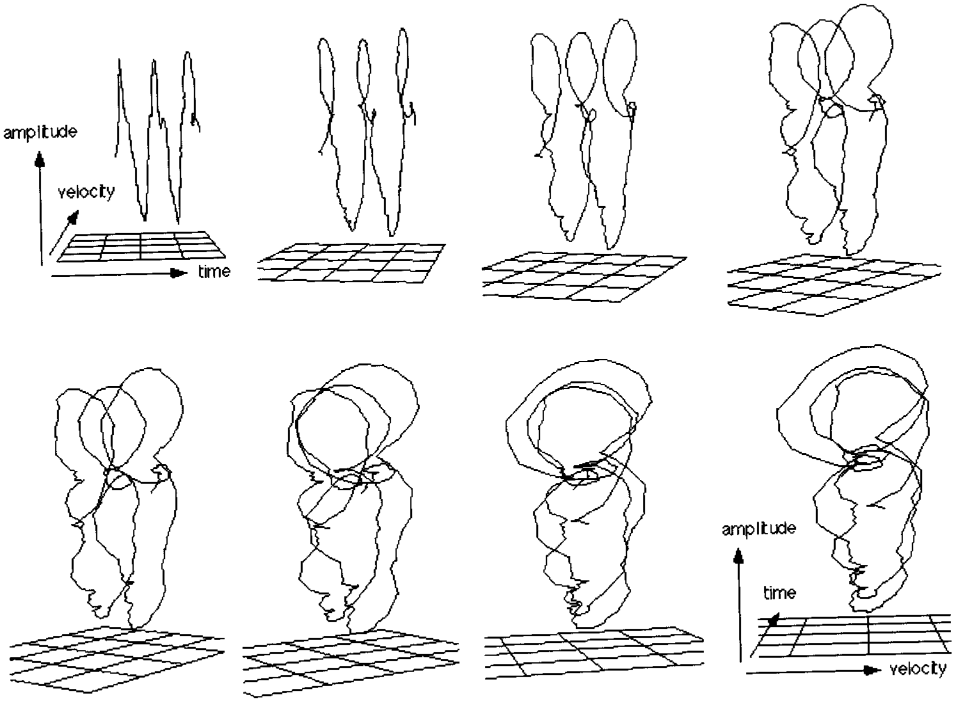

Fig. 2 A segment from the /a/ sound of "can't" in "clean.dip". The plots show that the wave spirals into an apparent centre and then spirals out again. The apparent centre is known as an "attractor" as the signal spirals in and as a "repellent" as it spirals out. Vowels generally show several of these so-called attractor/repellent basins.

3 PHASE PLOTS OF DIFFERENT TYPES OF SOUNDS

Acoustic speech signals are rather distinctive when viewed as phase plots[1]. In figs. 3 and 4, vowels, nasals, unvoiced fricatives, and voiced fricatives are illustrated. On the basis of vector field similarities with the behaviour of well investigated systems, certain dynamic origins of observed patterns may be suggested.

3.1 Vowels

Typical action. A vowel typically spirals towards an attractor and back out towards a larger cycle. There may be one or more attractor/repellent basins.

Presumed dynamics. The vowel phase portrait reflects to a large extent the sound pressure wave imparted by the vocal cords. The phase plot shows typical spring-like behaviour.

Attractors (spiralling inwards). At large amplitudes, this may reflect the progressive loss and damping of energy in the vocal cords.

Repellents (spiralling outwards). (1) Small outward motions may be facilitated by "inverse friction": (2) Larger outward motions may relate to the reinfusion of energy.

3.2 Nasals

Typical action. Similar action to vowels, except that there tend to be fewer attractor/repellent basins.

Presumed dynamics. As for vowels with respect to presumed vocal cord action. If the multiplicity of basins relates to friction and damping, the reduced number of basins in nasals may relate to reduced friction in the nasal vocal tract.

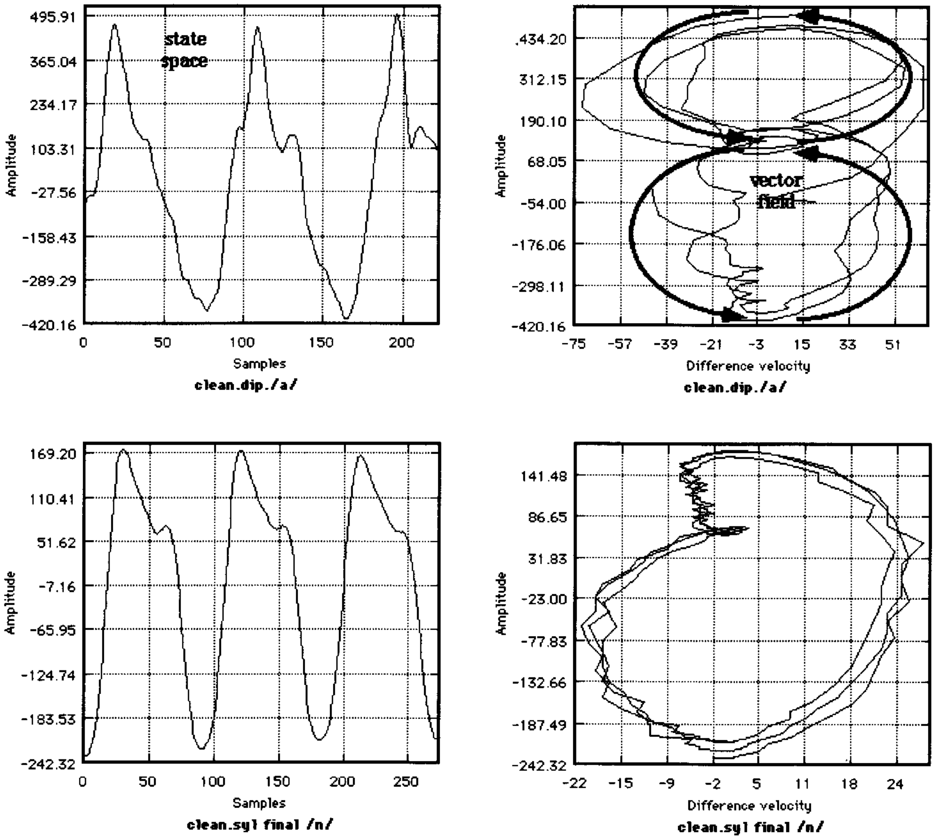

Fig. 3 Signal segments and corresponding phase portraits for different speech sounds.

3.3 Fricatives

Typical action. Strong phase excursions to the lower left and upper right are observed. No attractor/repellent basins are seen.

Presumed dynamics. Near instantaneous amplitude changes are associated with near-instantaneous velocity losses/gains. Voiced fricatives show a combination of the circular characteristics of voicing and the rapid oblique phase excursions characteristic of frication.

4 THE EFFECTS OF NOISE ('clean', 'dirty' and 'spont')

Many types of noise resemble frication. It may thus be possible to distinguish effects of noise from vocalic signals on the basis of their phase characteristics. Different levels of noise are illustrated below by small segments taken from the /a/ sounds found in the signals 'clean', 'dirty' and 'spont' (fig. 5). It can be seen that increased levels of noise are associated with a greater obscuring of vocalic phase characteristics: In noisy conditions, it becomes more difficult to identify attractor/repellent basins, and the outer extent of the phase portrait becomes harder to delineate.

However, a spline[2] operation on the signal prior to the derivation of the phase portrait helps identify vocalic aspects "hiding behind" the noise. In fig. 6, the vowel segments of fig. 5 are shown after interpolation with a tension spline set at a tension coefficient of 0.1.

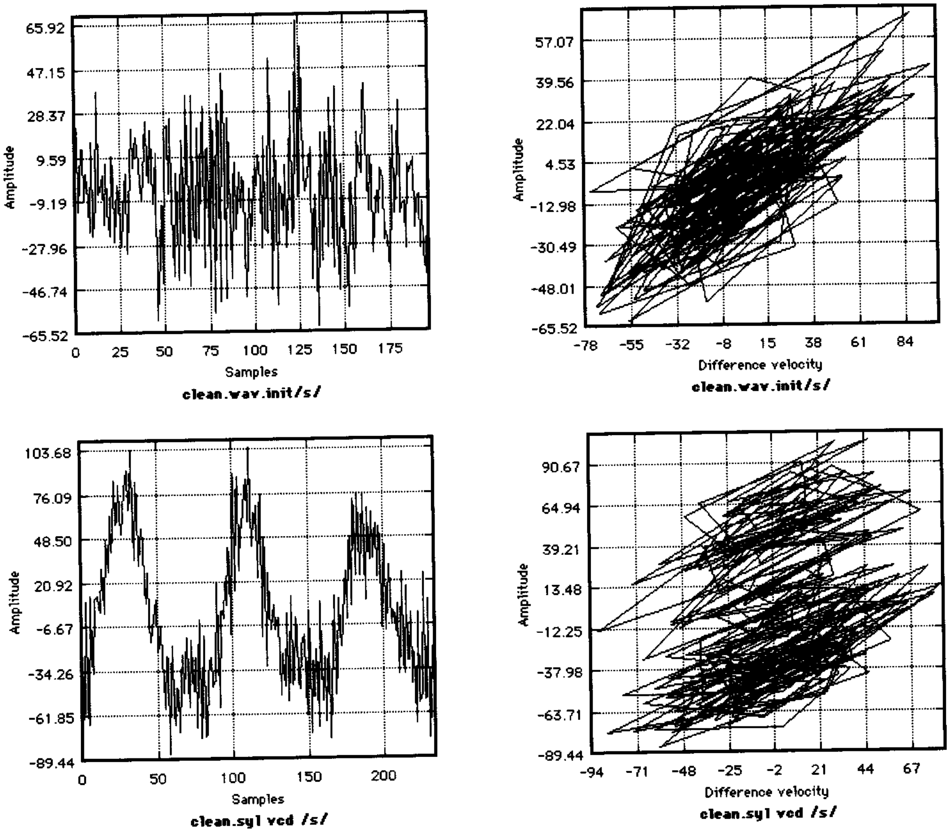

Fig. 4 Signal segments and corresponding phase portraits for different speech sounds.

In fig. 6 it can be readily seen that the splining operation helps reconstitute the vowel's low-frequency characteristics. A splined sound is not necessarily easier to perceive than the corresponding "noise-covered" sound, since high-frequency components are missing. However, splining appears to be a useful technique for rendering speech sounds more amenable to secondary vocalic analyses, such as pitch extraction.

5 CONCLUSION

Phase representations of speech sounds are useful for identifying the essential characteristics of different speech sounds. Vocalic and fricative types of sounds are readily distinguished on the basis of their phase characteristics. In noisy sounds, vocalic low-frequency phase characteristics can be reconstituted with a splining operation. This logic is likely to be of use in the voiced/unvoiced distinction required for pitch extraction algorithms. Further, the study of attractor/repellent basins (typical areas of loss and regaining of sound pressure energy) may provide useful insights into the speech sound's source-filter characteristics.

Notes

1 2D phase representations and splines used here are available in the Signalyze™ software (for availability, communicate with the author). The 3D representations will be part of a future version of Signalyze™. The program uses the phase characteristics described here to effect voiced/unvoiced distinctions in its Temporal Structure Analysis pitch extraction algorithm.

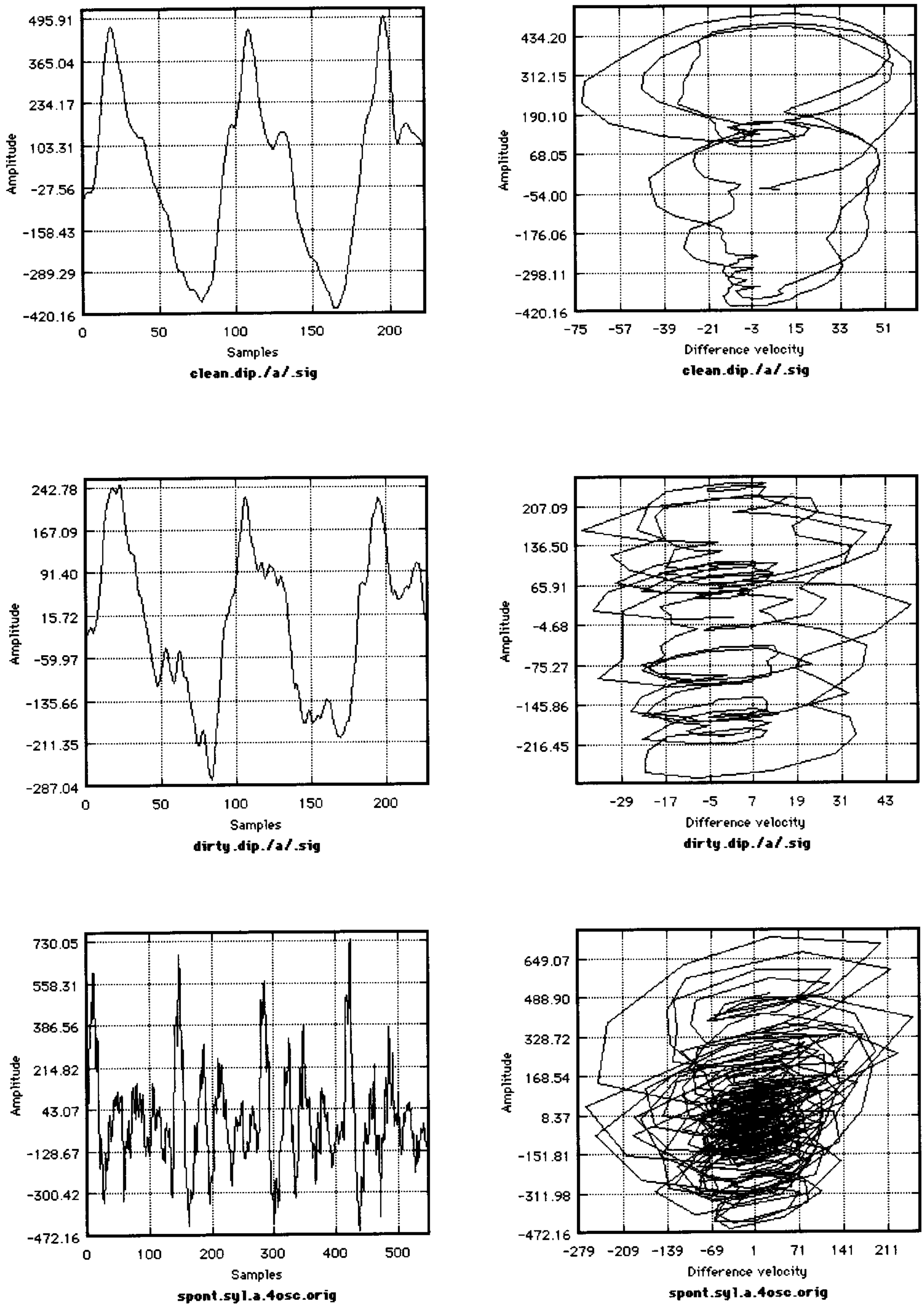

Fig. 5 Signal segments and corresponding phase portraits for the /a/ sound taken from 'clean', 'dirty' and 'spont'. Increasing levels of noise obscure the vowel's phase characteristics.

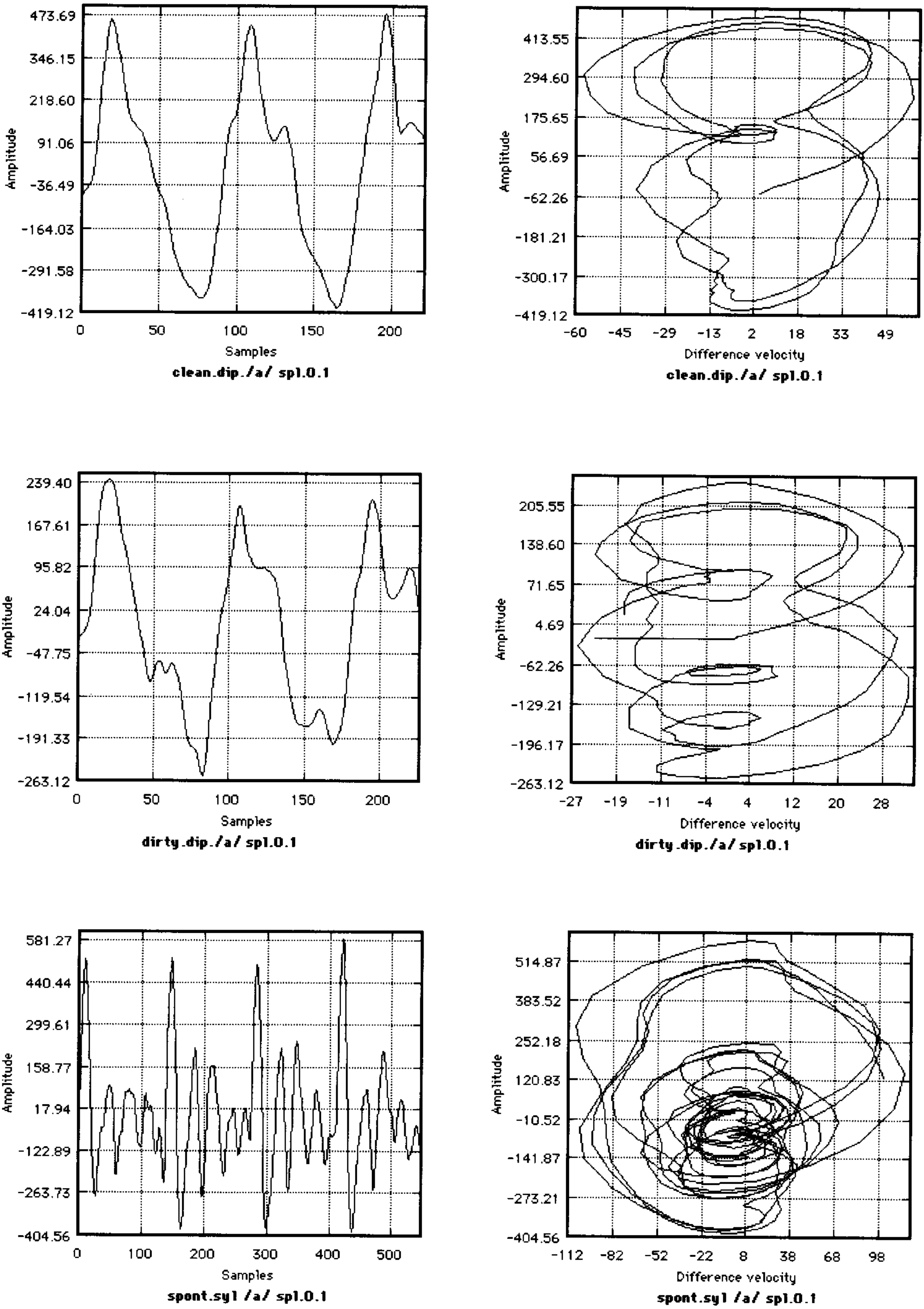

Fig. 6 The same signal segments and phase portraits for the /a/ sound taken from 'clean', 'dirty' and 'spont', but after interpolation with a tension spline, with a tension coefficient of 0.1. The vowel's low-frequency phase characteristics are reconstituted by the splining operation and the vowel's phase characteristics can be inspected.

2 Splines are cubic functions that "follow" a set of points. In a tension spline, the maximal tension is 1.0, which corresponds to a line that crosses the points exactly. Smaller coefficients cause the spline to approximate the points' positions, and to provide a smoother line.

REFERENCES

[1] R.H. Abraham & C.D. Shaw (1984), *The Visual Mathematics Library*, Vismath Volumes 0-4, Aerial Press Inc. Address: Aerial Press Inc. P.O. Box 1360, Santa Cruz, California 95061, USA.

[2] C. de Boor (1978), *A Practical Guide to Splines*, Springer-Verlag, Berlin.

[3] C. Demars (1990), *Représentation temps-fréquence et paramétrisations d'un signal: éléments de monographie*, Document LIMSI: 90-9, Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur, Orsay, France.

[4] J.A.S. Kelso & B. Tuller (1987), 'Intrinsic time in speech production: Theory, methodology, and preliminary observations', In: E. Keller and M. Gopnik (eds.), *Motor and Sensory Processes of Language*, Lawrence Erlbaum Associates Hillsdale, NJ, 203-222.

[5] R.M. Lerner (1959), *A Method of Speech Compression*, Ph.D. Thesis, MIT, Cambridge, MA.