Blind Inversion in Speech and Weapons Image Processing

John Hogden, CCS-3

The speech inversion problem—the problem of inferring the positions of the tongue, jaw, lips, and other *speech articulators* from a recording of a sound pressure wave over time—provides a nice example of a synergy between apparently disparate fields. Not only has speech inversion research benefited from using physical models of wave propagation, but processing techniques developed for speech inversion are starting to be applied to images used for hydrocode validation. We discuss an approach to the inversion problem, emphasizing aspects that generalize to other problems.

Research related to speech inversion includes work by former Los Alamos affiliates such as Kac (founding chairman of the Center for Nonlinear Studies External Advisory Committee) who asked whether it is possible to hear the shape of a drum [1], and Papcun et al. [2] who used a supervised algorithm to learn the mapping between simultaneously collected speech acoustics and articulator positions.

Our work extends the previous work by invoking a very general theorem to invert the mapping from articulator positions to speech acoustics. Since our method does not require simultaneous measurement of acoustics and articulation to learn the relationship, it is a type of blind inversion process. Ignoring caveats that are discussed in depth elsewhere [3], if we have a sufficiently long bandlimited signal, $\mathbf{x} = [x(1), x(2),...x(T)]$, that is difficult to observe (like articulator motions), and an observable signal (like speech acoustics), $\mathbf{y} = [y(1), y((2),...y(T)]$, that is a memoryless function of the unobservable signal, i.e., $y(t) = \mathrm{f}(x(t))$, then $\mathbf{y}$ can only have the same bandlimit as $\mathbf{x}$ if f() is an affine function, i.e., $y(t) = mx(t) + b$.

The theory stated above tells us that if we transform $\mathbf{y}$ by any g() such that $\hat{x}(t) = \mathrm{h}(x(t)) \equiv \mathrm{g}(f(x(t))) = \mathrm{g}(y(t))$, and $\hat{\mathbf{x}} = [\hat{x}(1), \hat{x}(2).....\hat{x}(T)]$ has the same bandlimit as $\mathbf{x}$, then h() will be affine. Thus, if we know the bandlimit on $\mathbf{x}$, we can recognize when we have found a g() that inverts f() to within an affine transform, and thereby get information about the unobservable signal.

In addition to proving the theorem stated above, we have developed an algorithm for finding g(). We have successfully tested our function inversion approach using speech data and computer generated, two-dimensional, bandlimited signals [4]. The left side of Fig. 1 shows periodic samples of the computer-generated signals. The right side of Fig. 1 shows the samples transformed by one of the nonlinearities we studied. Using our approach to find g() we were able to invert the nonlinear transformation. In this case, inverting the function requires reducing the dimensionality of the data, which is an important problem in its own right. We have also successfully applied the technique to synthetic data transformed by an f() with many-to-one mappings.

The theory does not require **x** and **y** to be related to speech. We can also think of *y(t)* as being the observable brightness of a pixel in an image and *x(t)* as being an unobservable value to be inferred from the image. By using appropriate physical constraints on **x**, our theory and closely related ideas have been used to help infer possible mix distributions in both real and simulated PINEX images [5].

For more information contact John Hogden at hogden@lanl.gov.

[1] M. Kac, *Am. Math. Monthly* **73** (4), Pt. 2, 1–23 (1966).
[2] G. Papcun, et al., *J. Acoustical Soc. Am.* **92** (2), Pt. 1, pp. 688–700 (1992).
[3] J. Hogden, et al., "Inverting Mappings from Smooth Paths through R^n to Paths through R^m: A Technique Applied to Recovering Articulation from Acoustics," *Speech Communication, Special Issue on Bridging the Gap Between Human and Automatic Speech Recognition*, Eds. K. Kirchhof & L. ten Bosch (Elsevier, 2007, http://dx.doi.org/doi:10.1016/j.specom.2007.02.008).
[4] J. Hogden, et al., "Blind Inversion of Multidimensional Functions for Speech Enhancement," in *Proceedings of the Eurospeech Conference* (2003), pp. 1409–1412.
[5] J. Hogden and R. Brewer, "Improvements on Blind Inversion of PINEX Data (U)," JOWOG-32M, Los Alamos, NM, September 26–30, 2005, Los Alamos National Laboratory report LA-CP-05-0990.

Funding Acknowledgements
NNSA's Advanced Simulation and Computing (ASC), Advanced Applications Program.

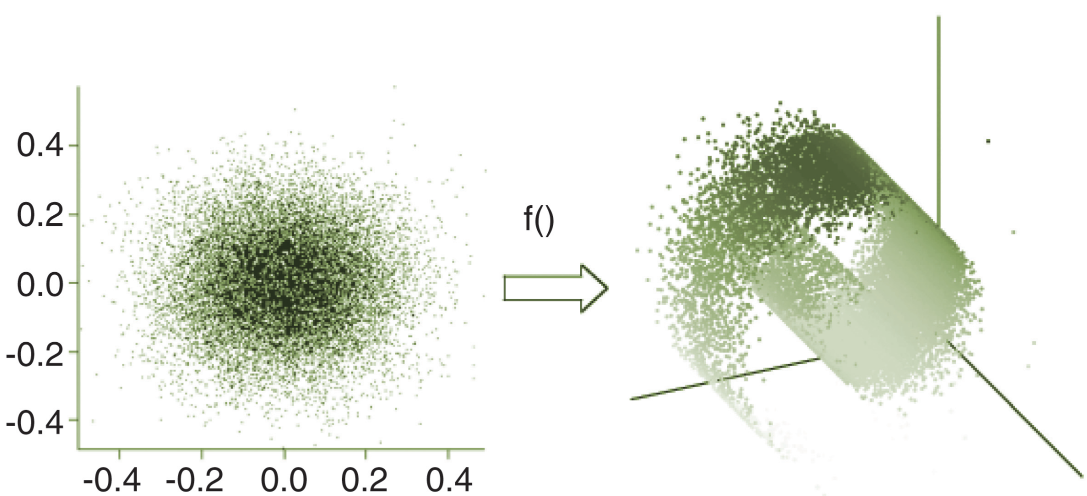

Fig. 1.
We did simulation work to see if we could blindly invert the f() shown above—a Cartesian-to-polar transformation followed by a swiss-roll transformation. We demonstrated that this and other complicated functions can be inverted by our technique.