A NEW ALGORITHM FOR INCORPORATING ACOUSTIC CONSTRAINTS INTO THE INVERSE SPEECH PROBLEM

James DeLucia and Fred Kochman

IDA Center for Communications Research
Thanet Road, Princeton, N.J. 08540-3699

ABSTRACT

We describe a new noniterative algorithm that generates the unique area function determined by the vocal tract length, the lip radius, and the spectral pair consisting of poles of the transfer function and zeros of the input impedance function. Our analysis is restricted to the class of piecewise-constant area functions defined on an even number of equal length intervals. The resulting algorithm involves fewer floating point operations per evaluation than the analogous method of Paige and Zue [4]. A method which uses a corpus of X-ray data is discussed for setting the higher order unobservable pole/zero frequencies.

1. INTRODUCTION

The general *inverse speech problem* is to determine the vocal tract shape from the speech signal. Considering only vowels, we let the area function $A(x)$ represent vocal tract shape. The first three formants $\{f_1, f_2, f_3\}$ are taken as the available acoustic information. The mapping from formants to areas is nonunique, so we consider the problem of generating area functions having a prescribed set of formants. We present a new algorithm for accomplishing this task in the significant special case of piecewise-constant area functions.

We model speech production by Webster's horn equation [7]

$$\frac{d}{dx}A(x)\frac{d}{dx}\Omega(x) + k^2 A(x)\Omega(x) = 0, \quad 0 < x < L \qquad (1)$$

where $k = 2\pi f/c$ (f being frequency and c the sound speed). The acoustic velocity $v(x) = \Omega'(x)$ (prime $'$ denotes d/dx) and pressure $p(x) = -\sqrt{-1}\rho ck\Omega(x)$ where ρ is the ambient mass density. Let $\{f_i\}$ and $\{g_i\}$ be the spectra of eigenfrequencies of (1) corresponding to the respective boundary conditions

$$\Omega'(0) = 0 \quad \text{and} \quad \begin{cases} \Omega(L) = 0 \Rightarrow \{f_i\} \\ \\ \Omega'(L) = 0 \Rightarrow \{g_i\} \end{cases} . \qquad (2)$$

The $\{f_i\}$ are poles of the transfer function (formants), $\{g_i\}$ are zeroes of the input impedance function (see [5]), and the pair $\{f_i, g_i\}$ are termed the *bi-spectrum*. Important to the inverse problem are the facts that zeroes are unobservable in the speech signal and that only the first three or four formants are predicted by (1).

Borg [1] proved that $\{f_i, g_i\}$ uniquely determines $A(x)$, up to a scale factor, for sufficiently regular coefficient functions. We have shown that an analogous result is true for piecewise constant (i.e. discontinuous) area functions, namely

$$A(x) = \sum_{n=1}^{N} A_n W\left(x - (n-1)\Delta\right), \qquad (3)$$

where $\Delta = L/N$ is the width of each interval and $W(x)$ is a *window function* which is unity when $0 \le x < \Delta$, but otherwise vanishes. We refer to $\vec{A} \equiv (A_1, A_2, \ldots, A_N)$ as the *area vector*, and only consider **even** $N = 2M$. In this case only M of the poles may be freely assigned and only $M-1$ of the zeros. This is a total of $N-1$ parameters, the same as the number of free parameters in the scaled area vector.

It becomes convenient to define the *area ratios vector*

$$\vec{r} \equiv \left(\frac{A_1}{A_2}, \ldots, \frac{A_{N-1}}{A_N}\right) \qquad (4)$$

and the *normalized truncated bispectrum vector*

$$\vec{h} \equiv (f_1, g_2, f_2, \ldots, g_M, f_M)/\hat{f} \qquad (5)$$

where $\hat{f} \equiv c/4L$. Note that both $\vec{r}$ and $\vec{h}$ have $N-1$ components. We refer to $\vec{h}$ simply as the *bispectrum vector* and label its elements

$$\left.\begin{aligned} h_{2m} &= f_m/\hat{f} \\ h_{2m-1} &= g_m/\hat{f} \end{aligned}\right\} \quad m = 1, \ldots, M \quad .$$

So $h_1 = 0$, $\vec{h} = (h_2, \ldots, h_N)$, and since poles and zeros are interlaced

$$0 < h_2 < \cdots < h_N < N. \qquad (6)$$

The vector $\vec{h}$ is a function via (1)-(3) of $\vec{r}$ alone. We formulate our results in terms of this function which we denote as the *forward map* and write as $\vec{h} = \mathcal{H}(\vec{r})$. A fairly explicit description of $\mathcal{H}$ is given in Theorem 1. Symmetries in the bispectrum are then described in Theorem 2. An important theorem then follows concerning the uniqueness of $\mathcal{H}$. After this we give an explicit computation of the *inverse map* $\vec{r} = \mathcal{H}^{-1}(\vec{h})$ in Theorem 4. Theorems 3 and 4 are the main results of this paper.

There is insufficient space available here to prove the theorems, however, the authors will supply them upon request.

Briefly, Theorem 1 is obtained by solving (1) for piecewise constant $A(x)$, as seen in (3), and requiring that boundary conditions (2) are satisfied. It is necessary to use the facts that both the acoustic pressure $p(x)$and volume velocity $A(x)v(x)$ are continuous functions of position. The first step of Theorem 2 relies on a technical construction in Sturm-Liouville theory known as the **Prüfer substitution** (See [9] chapter 10, sections 5-8), and the last two steps are obtained by analyzing the characteristic polynomials of Theorem 1. Theorems 3 and 4 are obtained by separately inverting and analyzing each step in the forward map (Theorem 1). Two matrices seen here are proven to be invertible by showing that they can be reduced to Vandermonde matrices.

Evaluating the inverse map is only part of a solution to the inverse vowel problem, since we are only given a few formants and not the complete vector $\vec{h}$. Even if we know L and a, before inverting we must fill in the remaining $\vec{h}$ values in such a way that the resulting area vector is anatomically meaningful. To this end, we explore in Section 3 a possible method of imposing static geometrical constraints into our model, thereby enabling us to set the higher order unobservable components of the bispectrum. Some discussion of setting the lower order unobservable zeroes, along with L and a is also seen in Section 3.

2. EVALUATING $\mathcal{H}$ AND $\mathcal{H}^{-1}$

The solution to (1)-(3) takes the form of a pair of *characteristic polynomials*, for each of which the zero-set is an eigenfrequency spectrum. Let $F(f)$ be the characteristic polynomial for poles and $G(f)$ the characteristic polynomial for zeroes. Following a few definitions we present in Theorem 1 the formulae for evaluating $F(f)$ and $G(f)$ for given f, $\vec{A}$ and L. Then Theorem 2 describes the relevant structure of the set of roots of $F(f)$ and $G(f)$.

Definition 1 *For $n = 1, \ldots, N-1$, let $r_n = A_n/A_{n+1}$, $p_n = (1+r_n)/2$, $q_n = (1-r_n)/2$, $\tilde{p}_n = (1+r_n^{-1})/2$, and $\tilde{q}_n = (1-r_n^{-1})/2$.*

Theorem 1 *For a tube of length L and area vector $\vec{A}$ having $N = 2M$ sections, the characteristic polynomials are given in terms of the angle $\theta = \pi f/(2M\hat{f})$ as*

$$F(f) = \sum_{m=0}^{M} \alpha_m \cos(m\theta)$$

and

$$G(f) = \frac{M\theta}{L} \sum_{m=1}^{M} \beta_m \sin(m\theta).$$

The coefficients $\{\alpha_m\}$ and $\{\beta_m\}$ are functions of $\vec{r}$ only. They are obtained in the following manner:

Let $\vec{U}^n \equiv (u_1^n, \ldots, u_n^n)$ be a real-valued n-dimensional row vector generated recursively by $\vec{U}^1 = (1)$ and, for $n = 1, \ldots, N-1$,

$$\vec{U}^{n+1} = p_n\,(u_1^n, \ldots, u_n^n, 0) + q_n\,(0, u_n^n, \ldots, u_1^n)\,. \qquad (7)$$

Then $\alpha_0 = u_{M+1}^N$, $\alpha_M = \beta_M = u_1^N$, and for $1 \le m < M$

$$\begin{bmatrix} \alpha_m \\ \beta_m \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} u_{M+1-m}^N \\ u_{M+1+m}^N \end{bmatrix} .$$

The set of roots has a structure which we will need to know when we construct a complete vector $\vec{h}$ given only the first few formants.

Theorem 2 *The bi-spectrum $\{f_i, g_i\}$ corresponding to a piecewise-constant tube of length L, having $N = 2M$ sections, has the following properties:*

1. *The values are real, they are* **interlaced**, *$g_1 = 0$ and $g_{M+1} = N\hat{f}$, independent of $\vec{A}$, i.e.*

$$0 < f_1 < g_2 < f_2 < \cdots < g_M < f_M < N\hat{f}.$$

2. *$\{f_1, \ldots, f_M\}$ determine $\{f_i\}$ by*
   - $f_{M+i} = -f_{M+1-i} + 2N\hat{f}, \quad 1 \le i \le M$
   - $f_i = f_{i-N} + 2N\hat{f}, \quad i > N$
3. *$\{g_2, \ldots, g_M\}$ determine $\{g_i\}$ by*
   - $g_{M+1+i} = -g_{M+1-i} + 2N\hat{f}, \quad 1 \le i < M$
   - $g_i = g_{i-N} + 2N\hat{f}, \quad i > N.$

Thus the M lowest order poles and $M-1$ lowest order zeroes (excluding g_1) determine the entire bi-spectrum.

So the forward problem is reduced to locating the first M positive roots of $F(f)$ and the first $M-1$ positive roots of $G(f)$, which is done numerically in practice. The functional dependence of the forward map upon $\vec{r}$ is apparent from Theorem 1, since the values A_n enter only through p_n and q_n, which depend only on r_n. The fact that $F(f)$ and $G(f)$ present themselves naturally as functions of $f/\hat{f}$ explains why we normalize in the definition (5) of $\vec{h}$. Theorem 2 shows that the full bispectrum contains no more information about $\vec{r}$ than does the normalized truncated bispectrum vector $\vec{h} = \mathcal{H}(\vec{r})$.

Theorem 3 *The map $\mathcal{H}$ is one-to-one. That is, if $\vec{h}$ can be obtained as $\vec{h} = \mathcal{H}(\vec{r})$, then $\vec{r}$ is unique.*

We now evaluate the inverse map for the class of piecewise-constant area functions by reversing the forward mapping algorithm of Theorem 1.

Theorem 4 *Consider a tube of length L composed of $N = 2M$ sections. If $\vec{h} = \mathcal{H}(\vec{r})$ then $\vec{r}$ can be recovered from $\vec{h}$ in the following five steps:*

1. *For all* $m = 1, \ldots, M$, *define the angles*

$$\theta_m = \frac{\pi}{2}\left(\frac{f_m}{M\hat{f}}\right) \quad \textit{and} \quad \phi_m = \frac{\pi}{2}\left(\frac{g_m}{M\hat{f}}\right).$$

2. *Construct the invertible matrices*

$$\mathbf{C} = \begin{bmatrix} 1 & \cos(\theta_1) & \cdots & \cos((M-1)\theta_1) \\ 1 & \cos(\theta_2) & \cdots & \cos((M-1)\theta_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos(\theta_M) & \cdots & \cos((M-1)\theta_M) \end{bmatrix}$$

and

$$\mathbf{S} = \begin{bmatrix} \sin(\phi_2) & \cdots & \sin((M-1)\phi_2) \\ \sin(\phi_3) & \cdots & \sin((M-1)\phi_3) \\ \vdots & \ddots & \vdots \\ \sin(\phi_M) & \cdots & \sin((M-1)\phi_M) \end{bmatrix}$$

along with the vectors

$$\vec{c} = \begin{bmatrix} \cos(M\theta_1) \\ \vdots \\ \cos(M\theta_M) \end{bmatrix} \quad \text{and} \quad \vec{s} = \begin{bmatrix} \sin(M\phi_2) \\ \vdots \\ \sin(M\phi_M) \end{bmatrix}.$$

3. Solve $\mathbf{C}\vec{\alpha} = -\vec{c}$ and $\mathbf{S}\vec{\beta} = -\vec{s}$ for $\vec{\alpha} = (\alpha_0, \ldots, \alpha_{M-1})^\dagger$ and $\vec{\beta} = (\beta_1, \ldots, \beta_{M-1})^\dagger$.
4. Generate $\vec{U}^N = \left(u_1^N, \ldots, u_N^N\right)$ from $\vec{\alpha}$ and $\vec{\beta}$ via

$$u_1^N = 1 \quad , \quad u_{M+1}^N = \alpha_0,$$

and for $1 \leq m < M$

$$\begin{bmatrix} u_{M+1-m}^N \\ \\ u_{M+1+m}^N \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \alpha_m \\ \\ \beta_m \end{bmatrix} .$$

5. Generate $\vec{r}$ from $\vec{U}^N$ by doing the following **backward recursion**: for $n = N-1, N-2, \ldots, 1$
   (a) $r_n = \left(u_1^{n+1} - u_{n+1}^{n+1}\right) / \left(u_1^{n+1} + u_{n+1}^{n+1}\right)$,
   (b) calculate $\tilde{p}_n$ and $\tilde{q}_n$ via Definitions 1,
   (c) generate $\vec{U}^n = (u_1^n, \ldots, u_n^n)$ via

$$\begin{aligned} \vec{U}^n \quad &= \left(u_1^{n+1}, \tilde{p}_n u_2^{n+1}, \ldots, \tilde{p}_n u_n^{n+1}\right) \\ &+ \left(u_{n+1}^{n+1}, \tilde{q}_n u_n^{n+1}, \ldots, \tilde{q}_n u_2^{n+1}\right). \end{aligned}$$

Note: While the vectors $\vec{\alpha}, \vec{\beta}$, *and each* $\vec{U}^n$ *calculated here may differ from their counterparts in Theorem 1 by a scale factor, the final result* $\vec{r}$ *is identical to the original.*

The algorithm of Theorem 4 should be compared with that of Paige and Zue [4] who considered the identical inverse speech problem. They adapted a method originally developed to analyze transmission lines to the problem of recovering the area vector from the bispectrum. Their algorithm is similar to ours in that it is noniterative and produces unique area vectors (up to scaling) from the tube length and the initial segments of pole/zero frequencies. However, ours is more efficient. The fractional increase in floating point operations of their algorithm as compared to ours is approximately $6/N$. This estimate is based on the use of Gaussian elimination to solve the M and $M-1$ dimensional sets of linear equations making up the first step in both algorithms. Faster matrix inversion routines will further increase this value. This cost difference occurs because their method requires doing two synthetic monomial divisions prior to estimating each component of the area vector.

3. STATIC CONSTRAINTS

To reconstruct an $\vec{A}$ which gives rise to formants $\{f_1, f_2, f_3\}$ do the following:

1. L and $\{f_1, f_2, f_3\} \Rightarrow h_{2m} = f_m/\hat{f}, \quad m = 1, 2, 3,$
2. choose $h_3, h_5, h_7, h_8, h_9, \ldots, h_N$ consistent with (6),
3. compute $\vec{r} = \mathcal{H}^{-1}(\vec{h})$ via Theorem 4,
4. $\vec{r}$ and $a \Rightarrow \vec{A}$ via $A_N = \pi a^2$ and $A_{n-1} = r_{n-1}A_n$.

Step 2) must be carried out in such a way that the resulting area vector is anatomically meaningful. Similarly a and L need to be chosen sensibly. Choosing these quantities will involve satisfying a combination of static and dynamic continuity constraints, a full investigation of which is beyond the scope of this paper. However, we consider below the use of static constraints to set the higher unknown components of the bispectrum.

In the manner of [8], X-ray measurements are used to incorporate anatomical constraints. Let $\{\vec{A}^1, \ldots, \vec{A}^J\}$ represent a corpus of area function measurements, and let $\vec{h}^j$ be the bispectrum vector corresponding to $\vec{A}^j$. We distill from this corpus the *average bispectrum vector* $(\mu_2, \ldots, \mu_N) \equiv J^{-1}\sum_{j=1}^{J}\vec{h}^j$. Eigenvalue perturbation studies [3, 6] suggest that the higher order components of $\vec{h}$ are of diminishing importance in distinguishing area functions, and might just as well be set to suitable nominal values. The idea discussed below is to always set unknown bispectrum components $(h_8, \ldots, h_N)$ to $(\mu_8, \ldots, \mu_N)$.

Yehia provided us with the area functions measurements used in [8]. All major vowel sounds are represented in this data base. We calculated $(\mu_2, \ldots, \mu_N)$ specific to this data and then replaced the upper components of each $\vec{h}^j$ by $(\mu_8, \ldots, \mu_N)$. The result of making these replacements are seen in Figure 1. Here $N = 20$ and cubic splines are used to produce smooth functions. The fine lines are the originals and the thick lines are the modified area functions. Clearly, using an X-ray corpus to set the upper components of $\vec{h}$ in this manner is a sensible technique. We feel that setting the lower unknown components (h_3, h_5, h_7), along with L and a, must be accomplished using dynamic continuity constraints. This is the subject of current research.

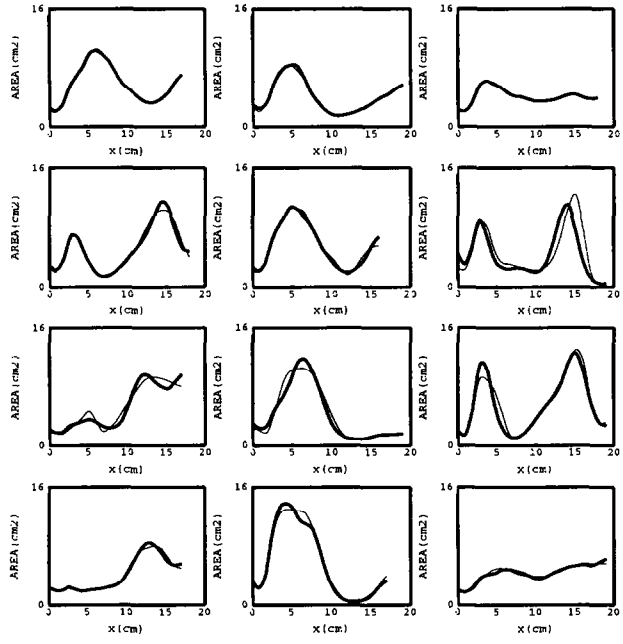

Figure 1. Replacing $(h_8, \ldots, h_N)$ with $(\mu_8, \ldots, \mu_N)$.

4. SUMMARY

For a piecewise constant area function, represented by $\vec{A}$, we have proven rigorously that the pole and zero spectra are completely determined by certain initial finite segments, and we have shown how to reconstruct $\vec{A}$, up to a scaling factor, given those initial segments and the tube length L. Except for the scaling factor, $\vec{A}$ is in fact unique. The algorithm for performing this unique reconstruction (Theorem 4) is new and is the main interest of this paper. It is seen to be more efficient than a previously developed algorithm [4] for solving the same problem.

We also considered the problem of incorporating anatomical constraints into the process of recovering $\vec{A}$ from formant information. We concluded that, given the tube length L, it is feasible to use a corpus of X-ray measurements to set the higher unknown components of the bispectrum vector.

REFERENCES

[1] G. Borg. Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe, *Acta Math.*, Vol. 78, 1945, pp.1-96.

[2] B. Gopinath and M. M. Sondhi. Determination of the Shape of the Human Vocal Tract form Acoustical Measurements, *The Bell System Technical Journal*, Vol. 49, No. 6, pp. 1195-1213, 1970.

[3] P. Mermelstein. Determination of the Vocal Tract Shape from Measured Formant Frequencies, *J. Acoust. Soc. Am.*, Vol. 41, No. 5, pp. 1283-1294, 1967.

[4] A. Paige and V. W. Zue. Computation of Vocal Tract Area Functions, *IEEE Transactions on Audio and Electroacoustics*, Vol. 18, No. 1, pp. 7-18, 1970.

[5] J. Schroeder and M. M. Sondhi. Techniques for Estimating Vocal-Tract Shapes from the Speech Signal, *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 1, Part II, pp. 133-150, 1994.

[6] M. R. Schroeder. Determination of the Geometry of the Human Vocal Tract by Acoustic Measurements, *J. Acoust. Soc. Am.*, Vol. 41, No. 4, pp. 1002-1010, 1966.

[7] A. G. Webster. Acoustical Impedance, and the Theory of Horns and of the Phonograph, *Proc. Natl. Acad. Sci. (U.S.)*, Vol. 5, pp. 275-282, 1919.

[8] Hani Yehia and Fumitada Itakura. A method to combine acoustic and morphological constraints in the speech production inverse problem, *Speech Communication*, Vol. 18, pp. 151-174, 1996.

[9] Garrett Birkhoff and Gian-Carlo Rota, *Ordinary Differential Equations*. (Blaisdell Publishing Company, 1962).