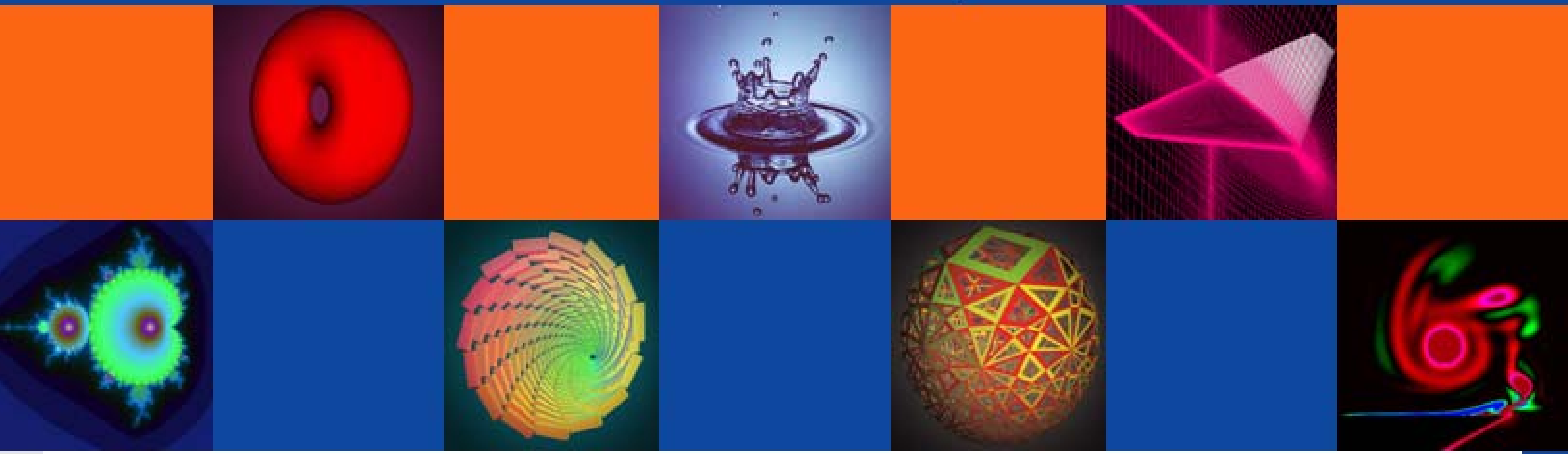
The University of Texas ARLINGTON

MATHEMATICS
Preprint Series



Determining the Shape of a Human Vocal Tract From Pressure Measurements at the Lips

Tuncay Aktosun

Technical Report 2007-01

http://www.uta.edu/math/preprint/

Determining the shape of a human vocal tract from pressure measurements at the lips

Tuncay Aktosun
Department of Mathematics
University of Texas at Arlington
Arlington, TX 76019-0408, USA

Abstract: In this review paper the sound propagation in a human vocal tract is analyzed. The sound pressure and the volume velocity in the tract are evaluated when a unit-amplitude, monochromatic, sinusoidal volume velocity is sent from the glottis towards the lips. The unique solution to the inverse problem is outlined, where the cross-sectional area of the vocal tract as a function of the distance from the glottis is recovered from the absolute value of the pressure at the lips known at all frequencies.

PACS (2003): 02.30.Zz, 43.72.Ct
Mathematics Subject Classification (2000): 34A55, 35R30, 76Q05
Keywords: Inverse scattering, Speech acoustics, Shape of vocal tract
Short title: Determining the shape of a vocal tract

1. INTRODUCTION

The production of human speech is similar in principle to the sound production in tubular musical instruments. Speech production can be visualized as follows: The inhaled air travels from the mouth down the vocal tract, a tube about 14-20 cm in length with a pair of lips at the mouth and another pair of lips known as the vocal cords at the opposite end. A flap known as the epiglottis blocks the passage to the esophagus, the muscular tube leading to the stomach, and the air passes between the vocal cords and travels to the lungs through another tube known as the voice box or the larynx. From the lungs the air travels back, and as it passes through the vocal cords, the opening between the vocal cords (known as the glottis) is adjusted depending on the sound produced. This is in some sense equivalent to controlling the cross sectional area of a water hose by pressing fingers on the hose at a particular location. The pressure created at the glottis puts the nearby air molecules in the vocal tract into longitudinal vibrations. These vibrations come out of the lips as a pressure wave, which is the sound created. The reality is certainly more complicated due to the presence of some articulators such as the tongue in the mouth. However, it can satisfactorily be assumed that by controlling the muscles around the vocal tract we are able to adjust the cross sectional area of the tract as a function of the distance from the glottis. Thanks to nonconstant cross sectional area functions, we are able to produce various sounds, each of which is unique to the individual speaker.

Human speech consists of basic units known as phonemes; for example, when we utter the word “book,” we in succession produce the three phonemes /b/, /u/, and /k/, each lasting about 10-20 msec. The number of phonemes depends on the language, and in fact the exact number of phonemes in a given language is often open to debate. For example, in American English one can argue that the number of phonemes is 36, 39, 42, or 45. The phonemes can be classified into two main groups as vowels and consonants. A vowel can be a monophthong such as /a/ as in “father” or a diphthong such as the middle sound in “goat.” The consonants can be classified into approximants (also called semivowels) such as /y/ in “yes,” fricatives such as /sh/ in “ship,” nasals such as /ng/ in “king,” plosives such as /p/ in “put,” and affricates such as /ch/ in “church.”

As far as a satisfactory mathematical description is concerned for vowel production, we can ignore the articulators and can assume that the cross sectional area of the vocal tract as a function of the distance from the glottis is the only factor responsible for the produced sound. During the production of each vowel the shape of the vocal tract can be assumed not to depend on time. In some sense this is analogous to watching a movie, where each frame contains a static image and lasts a short period; the continuous movie is observed and the continuous speech is heard as we watch a succession of static images and as we encounter a succession of static shapes of the vocal tract.

The air molecules at the glottal end can be put into vibration at various frequencies. The audible frequencies usually range from 20 Hz to 20000 Hz for human beings. We can assume that a vowel created at a particular frequency and the same vowel created at another frequency correspond to the same cross sectional area function, and hence we can use $A(x)$ to denote the cross sectional area of the vocal tract as a function of the distance x from the glottis. Having emphasized that $A(x)$ is independent of time and frequency during the production of each vowel, let us also mention that we can ignore the bending of the vocal tract and assume that the cross sectional area at each x value is circular with radius r, where

$$A(x) = \pi[r(x)]^2. \tag{1.1}$$

The pressure is the force per unit area felt as a result of the hits by the vibrating air molecules on that area. Let us use $p(x, t)$ to denote the pressure at the location x at time t, and let us note that the pressure produced at each location in the vocal tract, in particular at the lips, will certainly depend on the frequency. We expect the pressure to diminish as the frequency becomes zero; as the frequency increases from zero we can expect the pressure also to rise. At the other extreme, as the frequency becomes very large, we expect the pressure to settle down to an asymptotic value. The nearby air molecules vibrate in phase with each other at low frequencies and out of phase at very high frequencies; thus any further increases in frequency at large frequencies will not cause much additional increase in the pressure.

The effective velocity of all the air molecules crossing the area $A(x)$ will be denoted by $v(x,t)$, which has the dimension of speed times area, and this quantity is known as the volume velocity at the location x at time t. The length of the vocal tract depends on the individual speaker, and we will use l to denote that length. If the air vibrations in the vocal tract take place at a fixed frequency ν, then we can assume that the time dependence is sinusoidal and write

$$p(x,t) = P(k,x)\, e^{ikct}, \qquad v(x,t) = V(k,x)\, e^{ikct}, \tag{1.2}$$

where $i =: \sqrt{-1}$, c is the sound speed in the vocal tract, and k is the angular wavenumber measured in rad/sec and related to the frequency ν measured in Hz as

$$k = \frac{2\pi\nu}{c}.$$

The value of c can be treated as a constant, which is about 34300 cm/sec at room temperature. So, 20 Hz corresponds to $k = 0$ rad/cm and 20000 Hz corresponds to $k = 3.7$ rad/cm. The use of the complex exponential function is mathematically equivalent to but easier than using both the sine and cosine functions.

There are two main questions we can raise: For a given shape of the vocal tract, i.e. given $A(x)$ for $x \in (0,l)$, can we determine $P(k,x)$ and $V(k,x)$ for $k \in (0,+\infty)$ and $x \in (0,l)$? In particular, can we determine the absolute pressure $|P(k,l)|$ at the lips as a function of k when $A(x)$ is known for $x \in (0,l)$? This is known as a direct problem and can be described as determining the sound pressure at the lips as a function of frequency when the shape of the vocal tract is known. Conversely, can we determine $A(x)$ for $x \in (0,l)$ when we know $|P(k,l)|$ for $k \in (0,+\infty)$? This latter problem is known as an inverse problem, and it can be described as determining the shape of the vocal tract when the sound pressure is known at the lips at all frequencies. In this paper we present a brief review of these two problems and refer the reader to [1,2] and the references therein for the details.

We let $x = 0$ correspond to the glottis and $x = l$ to the lips. We can also assume that A is positive on $(0, l)$ and that both A and its derivative A' are continuous on $(0, l)$ and have finite limits at $x = 0^+$ and $x = l^-$. In fact, we can simply write $A(0)$ for the glottal area and $A(l)$ for the area of the opening at the lips because we will not use A or A' when $x \notin [0, l]$. Since the circular cross section as in (1.1) is known [11,12] to be a reasonable assumption, we can equivalently deal with quantities such as $r(x)$ and $r'(x)$ instead of $A(x)$ and $A'(x)$.

Besides t, x, l, c, k, $A(x)$, $r(x)$, $p(x, t)$, $v(x, t)$, $P(k, x)$, and $V(k, x)$, there is one more relevant constant, namely the air density μ (about 0.0012 gm/cm^3 at room temperature), and it is reasonable [11,12] to assume that the propagation is lossless and planar and that the acoustics in the vocal tract is governed [6,7,11-13] by

$$\begin{cases} A(x)\, p_x(x, t) + \mu\, v_t(x, t) = 0, \\ A(x)\, p_t(x, t) + c^2 \mu\, v_x(x, t) = 0, \end{cases} \tag{1.3}$$

where $t > 0$ and $x \in (0, l)$. In order to solve the direct problem uniquely, we need two side conditions, which we can choose by specifying the glottal volume velocity $v(0, t)$ and by assuming that the pressure wave at the lips is going out of the mouth only (i.e. there is no reflected pressure wave at the lips). For the former side condition it is convenient to use

$$v(0, t) = e^{ickt}, \qquad t > 0, \tag{1.4}$$

which corresponds to a unit-amplitude, sinusoidal volume velocity at the glottis at the angular wavenumber k. The latter condition is equivalent to rejecting $e^{ikl+ikct}$ and accepting only $e^{-ikl+ikct}$ in the expression for $p(x, t)$ when $x = l$.

Our review paper is organized as follows. In Section 2 we outline the solution to the direct problem and express $P(k, x)$ and $V(k, x)$ for $k \in (0, +\infty)$ and $x \in (0, l)$ appearing in (1.2) in terms of the relevant quantities c, μ, l, k, $A(x)$, and the so-called Jost solution $f(k, x)$ to a related Schrödinger equation; in that section we also indicate how $f(k, x)$ is obtained from l, k, and $A(x)$. In Section 3 we outline the solution to the inverse problem and show how $A(x)$ for $x \in (0, l)$ can uniquely be obtained from $|P(k, l)|$ known for $k \in (0, +\infty)$.

2. SOLUTION TO THE DIRECT PROBLEM

In this section we relate the acoustic system in (1.3) to the Schrödinger equation, and we present the explicit expressions for the pressure and the volume velocity in the vocal tract compatible with the aforementioned two side conditions.

Using (1.1) and (1.2) we can write (1.3) as

$$\begin{cases} \pi r(x)^2 P'(k,x) + ic\mu k\, V(k,x) = 0, \\ c^2 \mu\, V'(k,x) + i\pi ck\, r(x)^2 P(k,x) = 0, \end{cases} \tag{2.1}$$

where the prime denotes the-x derivative. Eliminating V in (2.1), we get

$$[r(x)^2\, P'(k,x)]' + k^2 r(x)^2 P(k,x) = 0, \qquad x \in (0,l), \tag{2.2}$$

or eliminating P we get

$$\left[\frac{V'(k,x)}{r(x)^2}\right]' + k^2\, \frac{V(k,x)}{r(x)^2} = 0, \qquad x \in (0,l). \tag{2.3}$$

With the help of (1.2) and (2.1) we see that (1.4) is equivalent to

$$P'(k,0) = -\frac{ic\mu k}{\pi r(0)^2}, \quad V(k,0) = 1, \tag{2.4}$$

and the condition that the pressure wave at the lips is only traveling out of and not into the mouth is equivalent to

$$P'(k,l) = -\left[ik + \frac{r'(l)}{r(l)}\right] P(k,l), \quad k^2 V(k,l) = \left[ik + \frac{r'(l)}{r(l)}\right] V'(k,l). \tag{2.5}$$

By letting

$$\psi(k,x) =: r(x)\, P(k,x),$$

we can transform (2.2) into the Schrödinger equation

$$\psi''(k,x) + k^2 \psi(k,x) = \frac{r''(x)}{r(x)}\, \psi(k,x), \qquad x \in (0,l), \tag{2.6}$$

where $r''(x)/r(x)$, the relative concavity of the radius, acts as a potential. In order to analyze the direct and inverse problems associated with the vocal tract, we will be using two particular solutions to (2.6), one of which is the Jost solution $f(k,x)$ satisfying the initial conditions

$$f(k,l)=e^{ikl},\quad f'(k,l)=ik\,e^{ikl}, \tag{2.7}$$

and the other is the regular solution $\varphi(k,x)$ satisfying the initial conditions

$$\varphi(k,0)=1,\quad \varphi'(k,0)=\frac{r'(0)}{r(0)}. \tag{2.8}$$

It is known that [3,5,9,10] $f(-k,x)$ is also a solution to (2.6) and that, for each fixed nonzero k, the functions $f(k,x)$ and $f(-k,x)$ are linearly independent. Let us also note that from (2.6) and (2.8) it follows that

$$\varphi(0,x)=\frac{r(x)}{r(0)},\qquad x\in(0,l), \tag{2.9}$$

which will play a key role in solving our inverse problem. We define the Jost function $F(k)$ as

$$F(k):=-i\left[f'(k,0)-\frac{r'(0)}{r(0)}\,f(k,0)\right], \tag{2.10}$$

which will be useful in our analysis. It is known that [4,8-10]

$$F(k)=k+O(1),\qquad k\to+\infty, \tag{2.11}$$

$$F(-k)=-F_\alpha(k)^*,\qquad k\in(-\infty,+\infty), \tag{2.12}$$

where the asterisk denotes complex conjugation.

The proof of the following theorem is straightforward.

Theorem *The acoustic equation (2.2) with the boundary conditions given in the first formulas in (2.4) and (2.5) has a unique solution, which is given by*

$$P(k,x)=-\frac{c\mu k\,f(-k,x)}{\pi\,r(0)\,r(x)\,F(-k)},\qquad x\in(0,l), \tag{2.13}$$

where $f(k,x)$ is the Jost solution to (2.6) and $F(k)$ is the Jost function defined in (2.10). Similarly, (2.3) with the boundary conditions given in the second formulas in (2.4) and (2.5) is uniquely solvable, and the solution is given by

$$V(k,x)=-\frac{i\,r(x)}{r(0)\,F(-k)}\left[f'(-k,x)-\frac{r'(x)}{r(x)}\,f(-k,x)\right],\qquad x\in(0,l).$$

Using (2.7) and (2.12) in (2.13) the absolute pressure at the lips is seen to be

$$|P(k,l)| = \frac{c\mu k}{\pi\, r(0)\, r(l)\, |F(k)|}, \qquad k \in (0,+\infty). \tag{2.14}$$

Note that, by using (2.11) in (2.14), we get

$$P_\infty := \lim_{k\to+\infty} |P(k,l)| = \frac{c\mu}{\pi\, r(0)\, r(l)}, \tag{2.15}$$

and hence P_∞, the asymptotic value of the absolute pressure, is uniquely determined by the cross sectional areas at the glottis and at the lips. From (2.14) and (2.15) we see that

$$\frac{|P(k,l)|}{P_\infty} = \frac{k}{|F(k)|}, \qquad k \in (0,+\infty), \tag{2.16}$$

which will play an important role in our analysis of the inverse problem via the Gel'fand-Levitan method in the next section.

3. SOLUTION TO THE INVERSE PROBLEM

In this section we show that the cross sectional area of the vocal tract as a function of the distance from the glottis can uniquely be recovered from the absolute pressure at the lips known at all frequencies. This inverse problem can be solved by various methods [1,2], and here we outline the solution with the help of the Gel'fand-Levitan method [4,8-10].

a) First, obtain the value of P_∞ from the data $|P(k,l)|$ known for $k \in (0,+\infty)$. In practice, it turns out that $|P(k,l)|$ reaches its asymptotic value in the audible range, i.e. somewhere with $k \in (0, 3.7)$, where we recall that $k = 3.7$ rad/cm corresponds to 20000 Hz.

b) By using (2.16), form the Gel'fand-Levitan kernel $G(x,y)$, which is defined as

$$G(x,y) := \frac{2}{\pi} \int_0^\infty dk \left[\frac{|P(k,l)|^2}{P_\infty^2} - 1 \right] \cos(kx)\, \cos(ky).$$

c) It is known that the Gel'fand-Levitan integral equation

$$h(x,y) + G(x,y) + \int_0^x dz\, G(y,z)\, h(x,z) = 0, \qquad 0 \le y < x,$$

is uniquely solvable, and its solution $h(x,y)$ yields

$$\varphi(k,x) = \cos kx + \int_0^x dy\, h(x,y)\, \cos ky, \qquad x \in (0,l), \tag{3.1}$$

where $\varphi(k,x)$ is the regular solution appearing in (2.8) and (2.9).

d) From (2.9), (2.15), and (3.1), obtain

$$r(0) = \sqrt{\frac{c\mu}{\pi\, P_\infty\, \varphi(0,l)}}, \tag{3.2}$$

and hence the cross sectional area of the vocal tract as a function of the distance from the glottis is obtained by using (3.1) and (3.2) in (2.9) as

$$A(x) = \pi\, r(0)^2 \varphi(0,x)^2,$$

or equivalently as

$$A(x) = \frac{c\mu \left[1 + \int_0^x dy\, h(x,y)\right]^2}{P_\infty \left[1 + \int_0^l dy\, h(l,y)\right]}.$$

Acknowledgments. The research leading to this article was supported in part by the National Science Foundation under grant DMS-0610494. The author is grateful to the colleagues in the Department of Applied Mathematics at Changwon National University for their hospitality during the fifth Changwon International Symposium on Advanced Science and Technology in 2006.

REFERENCES

[1] T. Aktosun, *Inverse scattering for vowel articulation with frequency-domain data,* Inverse Problems **21**, 899–914 (2005).

[2] T. Aktosun, *Inverse scattering to determine the shape of a vocal tract,* in: M. A. Dritschel (ed.), *The extended field of operator theory,* Birkhäuser, Basel, 2006, to appear.

[3] T. Aktosun and M. Klaus, *Chapter 2.2.4, Inverse theory: problem on the line,* in: E. R. Pike and P. C. Sabatier (eds.), *Scattering,* Academic Press, London, 2001, pp. 770–785.

[4] T. Aktosun, and R. Weder, *Inverse spectral-scattering problem with two sets of discrete spectra for the radial Schrödinger equation,* Inverse Problems **22**, 89–114 (2006).

[5] K. Chadan and P. C. Sabatier, *Inverse problems in quantum scattering theory,* 2nd ed., Springer, New York, 1989.

[6] G. Fant, *Acoustic theory of speech production,* Mouton, The Hague, 1970.

[7] J. L. Flanagan, *Speech analysis synthesis and perception,* 2nd ed., Springer, New York, 1972.

[8] I. M. Gel'fand and B. M. Levitan, *On the determination of a differential equation from its spectral function,* Am. Math. Soc. Transl. (ser. 2) **1**, 253–304 (1955).

[9] B. M. Levitan, *Inverse Sturm-Liouville problems,* VNU Science Press, Utrecht, 1987.

[10] V. A. Marchenko, *Sturm-Liouville operators and applications,* Birkhäuser, Basel, 1986.

[11] J. Schroeter and M. M. Sondhi, *Techniques for estimating vocal-tract shapes from the speech signal,* IEEE Trans. Speech Audio Process. **2**, 133–149 (1994).

[12] M. M. Sondhi, *A survey of the vocal tract inverse problem: theory, computations and experiments,* in: F. Santosa, Y. H. Pao, W. W. Symes, and C. Holland (eds.), *Inverse problems of acoustic and elastic waves,* SIAM, Philadelphia, 1984, pp. 1–19.

[13] K. N. Stevens, *Acoustic phonetics,* MIT Press, Cambridge, MA, 1998.