Area function from acoustic measurements with articulatory constraints:

Historical perspective and geometric constructions,

Paul Milenkovic,
University of Wisconsin-Madison,
milenkovic@engr.wisc.edu

presentation to Avaya Labs,
December 3, 2001.

Goal

Determine the vocal tract area function from measurements of the acoustic speech waveform (formant frequencies).

Application

Speech synthesis with coupled voice source,

Synthesizing CV's with acoustic/aerodynamic interaction,

Research/clinical measurement of articulation,

Speech recognition in articulatory space.

Status

Fant-Stevens nomograms used to inform general high/low, front/back distinctions from spectrograms,

More specific determinations of area function have not found widespread use.

Outline

Levinson-Durbin algorithm and system ID of layered media,

Model for vocal tract autocorrelation function and illustration of many-to-one articulatory-to-acoustic mapping,

Sine-cosine 2-basis model for area function – unifies Harshman and Ladefoged factor model with Stevens-Fant tongue constriction model,

Broken-line construction of tongue outline in midsagittal plane by displacement from palate outline,

Tongue model with local displacements, tongue tip articulation,

Gridding the 2-D vocal tract by constructing circles that fill the space between palate and tongue,

Refining the grid by graphical approximation to conformal mapping.

Levinson-Durbin algorithm for identification of autoregressive systems and connection to layered media:

$$\delta[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow h[n]$$

$$r[n] = h[n] * h[-n]$$

$$h[n] \longrightarrow \boxed{\hat{A}(z)} \longrightarrow e[n]$$

$$E = \sum e^2[n] = \text{MINIMUM } (=1)$$

WHEN $\tilde{A}(z) = A(z)$ FOR

$$A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p}$$

$$\hat{A}(z) = 1 + \hat{a}_1 z^{-1} + \ldots + \hat{a}_p z^{-p}$$

$$a_0 = \tilde{a}_0 = 1$$

$$e_p[n] = h[n] + [a_1, \ldots, a_p] \begin{bmatrix} — h[n-1] — \\ \vdots \\ — h[n-p] — \end{bmatrix}$$

$$= h[n] - [k_1, \ldots, k_p] \begin{bmatrix} — f_0[n] — \\ — f_{p-1}[n] — \end{bmatrix}$$

PARTIAL CORRELATIONS

ORTHOGONAL BASIS

WHERE $$k_j = \frac{\langle h[n], f_{j-1}[n] \rangle}{\langle f_{j-1}[n], f_{j-1}[n] \rangle}$$

0 through order p-1 backward prediction residuals form an orthonormal basis (least-squares orthogonality principle).

Backward predictors in turn are forward predictors turned around (from stationarity of r[n,m] = r[n-m]).

FORWARD PREDICTOR

$$e_{p-1}[n] = h[n] + a_1^{p-1} h[n-1] + \ldots + a_{p-1}^{p-1} h[n-p+1]$$

BACKWARD PREDICTOR

$$f_{p-1}[n] = \quad a_{p-1}^{p-1} h[n-1] + \ldots + a_1^{p-1} h[n-p+1] + h[n-p]$$

PARTIAL CORRELATION

$$k_p = \frac{\langle h[n], f_{p-1}[n] \rangle}{\langle f_{p-1}[n], f_{p-1}[n] \rangle}$$

$$= \frac{a_{p-1}^{p-1} r[1] + \ldots + a_1^{p-1} r[p-1] + r[p]}{E_{p-1}}$$

ORDER UPDATE

$$e_p[n] = e_{p-1}[n] - k_p f_{p-1}[n]$$

SO

$$a_1^p = a_1^{p-1} - k_p a_{p-1}^{p-1}$$

$$\vdots$$

$$a_{p-1}^p = a_{p-1}^{p-1} - k_p a_1^{p-1}$$

$$a_p^p = -k_p$$

AND

$$\langle e_{p-1}[n], e_{p-1}[n] \rangle = \langle f_{p-1}[n], f_{p-1}[n] \rangle = E_{p-1}$$

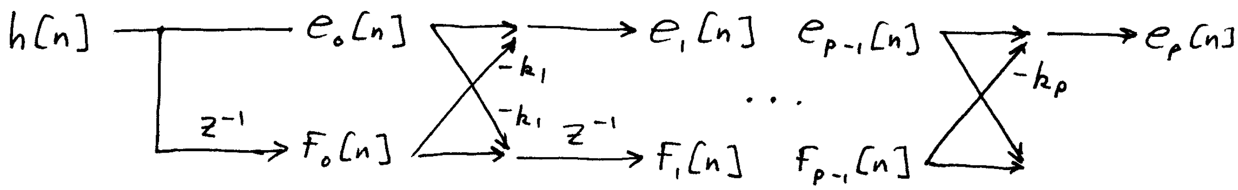
$$\langle e_{p-1}[n], f_{p-1}[n] \rangle = 0$$

SO

$$E_p = (1 - k_{p-1}^2) E_{p-1}^2 \qquad E_0 = r[0]$$

INVERSE LATTICE

$$\left(h[n] \longrightarrow \boxed{A(z)} \longrightarrow s[n] \right)$$

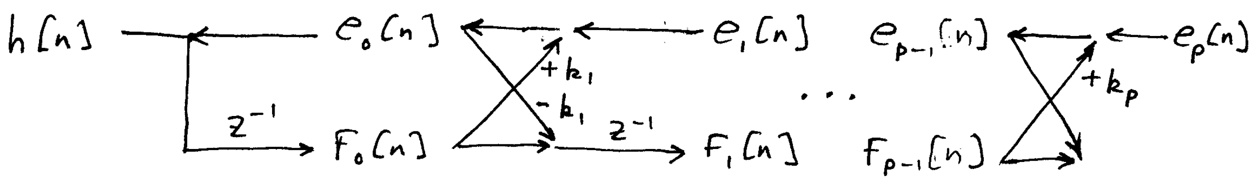


NOTE: $e_1[n] = e_0[n] - k_1 f_0[n]$

OR $e_0[n] = e_1[n] + k_1 f_0[n]$

FORWARD LATTICE

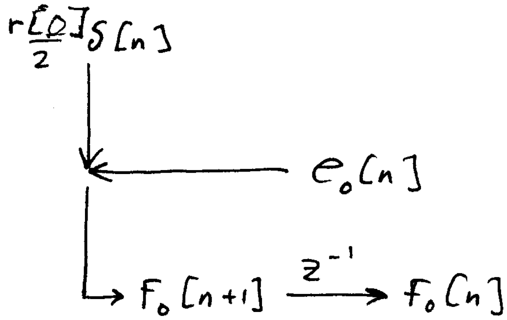
$$\left(s[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow h[n] \right)$$



The inverse lattice is simply an expression of the forward-predictor order update. The forward lattice is the exact same signal flow graph with the directions of the upper path turned around.

The connection between these lattices and acoustic tubes is left as an exercise for the reader.

Output-terminal impulse response of layered system with reflecting cap layer:



$$h_T[n] = e_0[n] + F_0[n+1]$$
$$= \tfrac{1}{2} r[0] \qquad n=0$$
$$= 2\, e_0[n] \qquad n>0$$

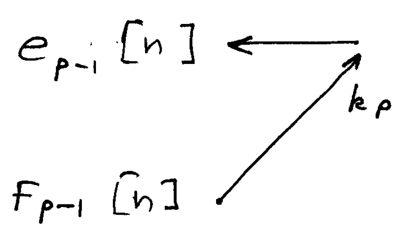
FURTHERMORE

$$h_T[n] \longrightarrow \boxed{A_{p-1}(z)} \longrightarrow 2\, e_{p-1}[n]$$

FOR $n \geq p$

BECAUSE $h_T[n]$ IS UNFORCED FOR $n \geq 1$

P'TH LAYER



IN REALITY, P IS ROUND TRIP, BUT WE COLLECT TOLLS LIKE THE PORT AUTHORITY

$$k_p = \frac{e_{p-1}[p]}{F_{p-1}[p]}$$

TAKES $n = p$ TIME FOR IMPULSE TO PENETRATE TO P'TH LAYER

$$= \frac{\frac{1}{2}\left\{ h_T[p] + a_1^{p-1} h_T[p-1] + \ldots + a_{p-1}^{p-1} h_T[1] \right\}}{\frac{1}{2} r[0]\left(1-k_1^2\right)\cdots\left(1-k_{p-1}^2\right)}$$

NEW JERSEY STYLE ROUND-ABOUTS AS WAVE PASSES LAYERS

Conclusion:

The vocal tract terminal impulse response is the causal part of the symmetric vocal tract transfer-function autocorrelation function according to

$$h_T[0] = \frac{1}{2} r[0]$$

$$h_T[n] = r[n] \qquad n \geq 0$$

and Levinson-Durbin determines layered structure from terminal impulse response. See A. Bruckstein and T. Kailath (1987), An inverse scattering framework for several problems in signal processing, IEEE ASSP Magazine 4, 6-20.

Hilbert relations say magnitude-squared transfer function in the real part of the terminal impedance, establishing correspondence between transfer function poles (formants) and poles of the terminal impedance.

Model for vocal tract autocorrelation function r[n] that generates the many-to-one articulatory-to-acoustic mapping:

TRANSFER FUNCTION

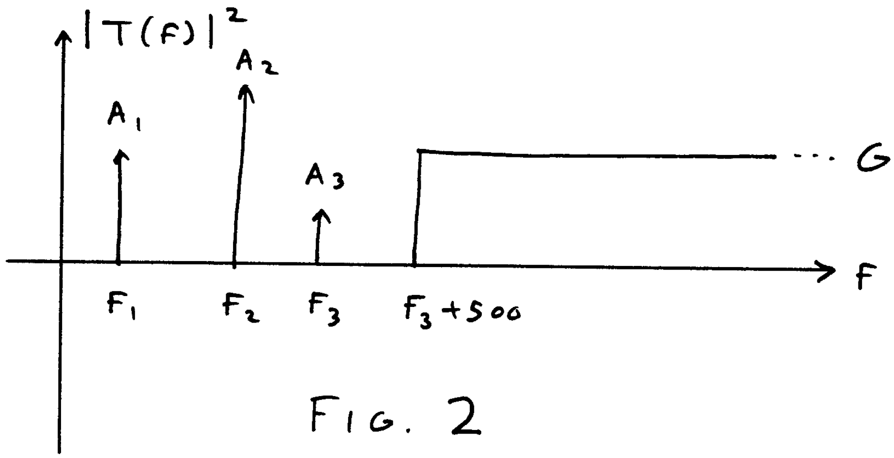


FIG. 2

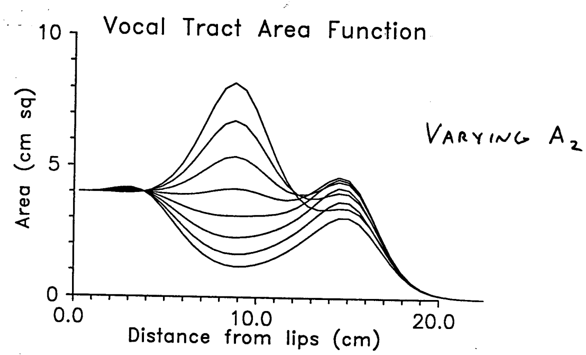
AUTOCORRELATION FUNCTION

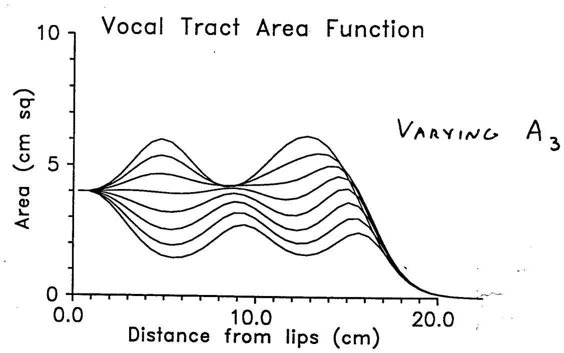
$$
\begin{aligned}
r[m] = \; & A_1 \cos 2\pi F_1 mT \; + \\
& A_2 \cos 2\pi F_2 mT \; + \\
& A_3 \cos 2\pi F_3 mT \; + \\
& G \, h_{HP}(mT)
\end{aligned}
$$

F_1, F_2, F_3 : KNOWN MODE (FORMANT) FREQUENCIES

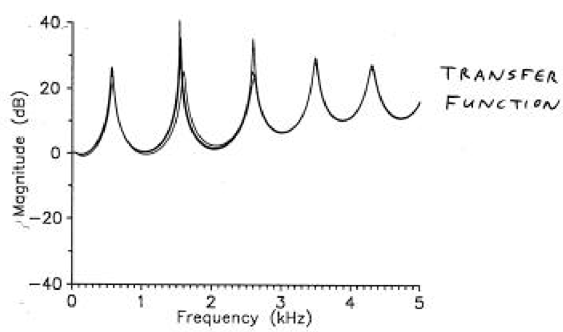
A_1 : CONTROLS $r[0]$

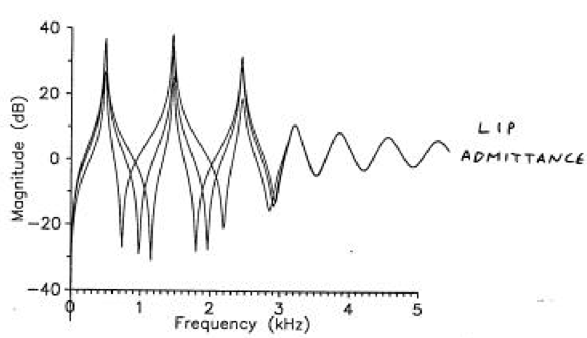
A_2, A_3, G : CONTROLS AREA FUNCTION


Vocal Tract Area Function
10
5
0
Area (cm sq)
0.0
10.0
20.0
Distance from lips (cm)
VARYING A2

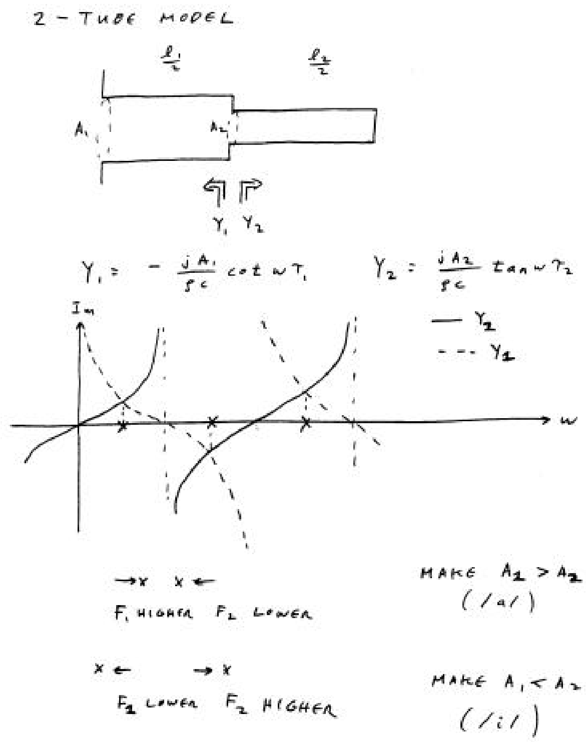

Vocal Tract Area Function
10
5
0
Area (cm sq)
0.0
10.0
20.0
Distance from lips (cm)
VARYING A3

VARYING A2

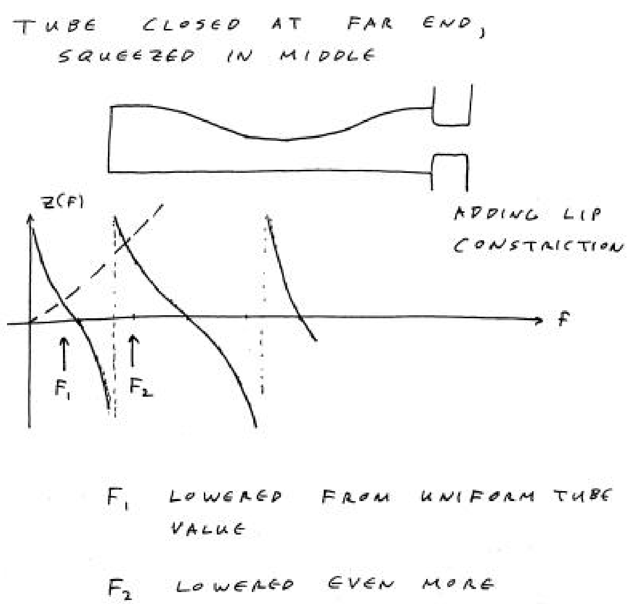

40
20
0
−20
−40
Magnitude (dB)
0
1
2
3
4
5
Frequency (kHz)
TRANSFER FUNCTION


40
20
0
−20
−40
Magnitude (dB)
0
1
2
3
4
5
Frequency (kHz)
LIP ADMITTANCE

Symmetrical deformation of acoustic tube moves terminal impedance zeroes while leaving formants fixed. Asymmetrical deformation moves formants:



So, how do we lower F2 to get /u/?

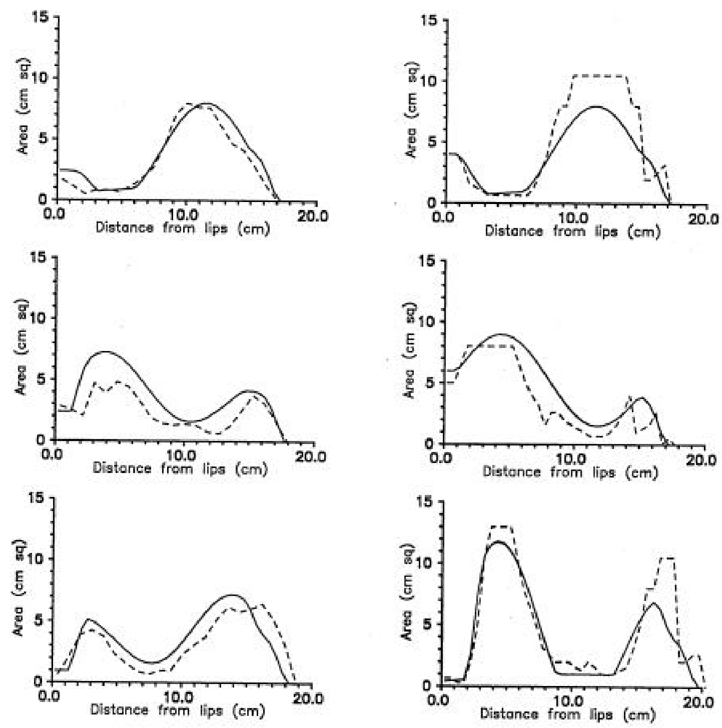


A sine-cosine 2-basis model

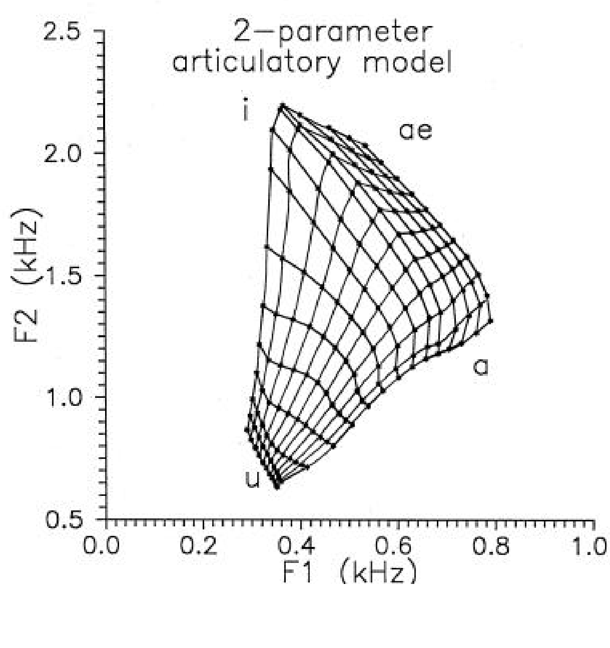
Unifies factor (Harshman-Ladefoged) model with

Tongue constriction (Fant-Stevents) model

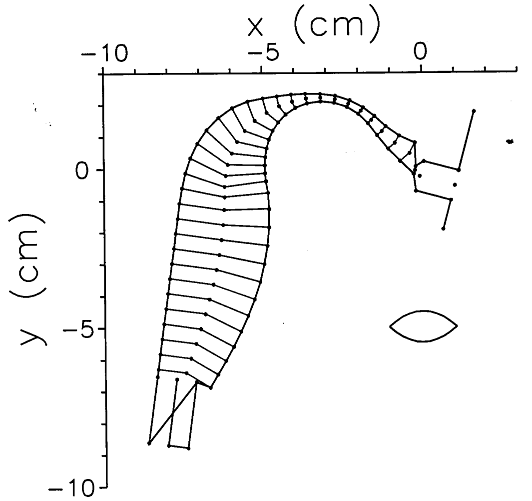
This model can represent the geometric space of vowels

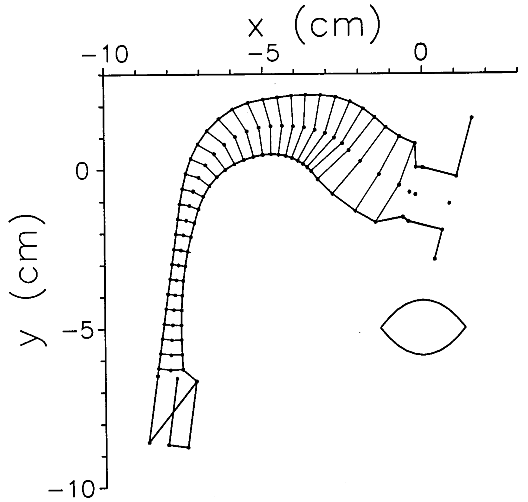


and with control of the lips, the acoustic space of vowels

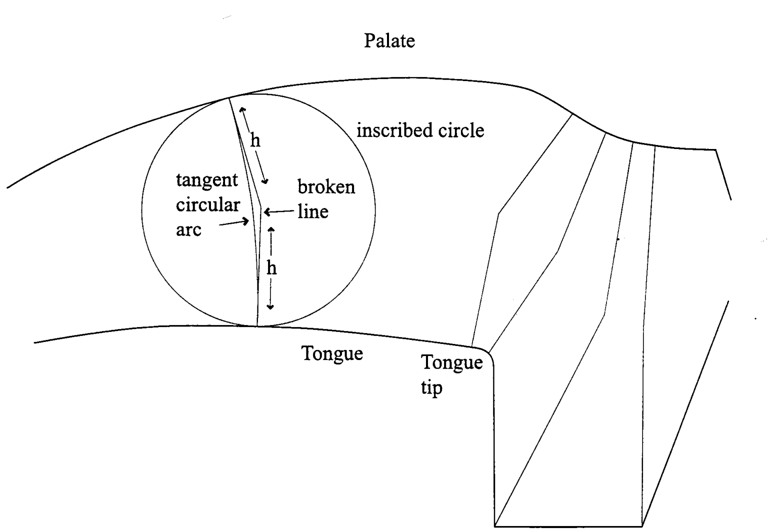


Midsagittal representation of 2-basis model

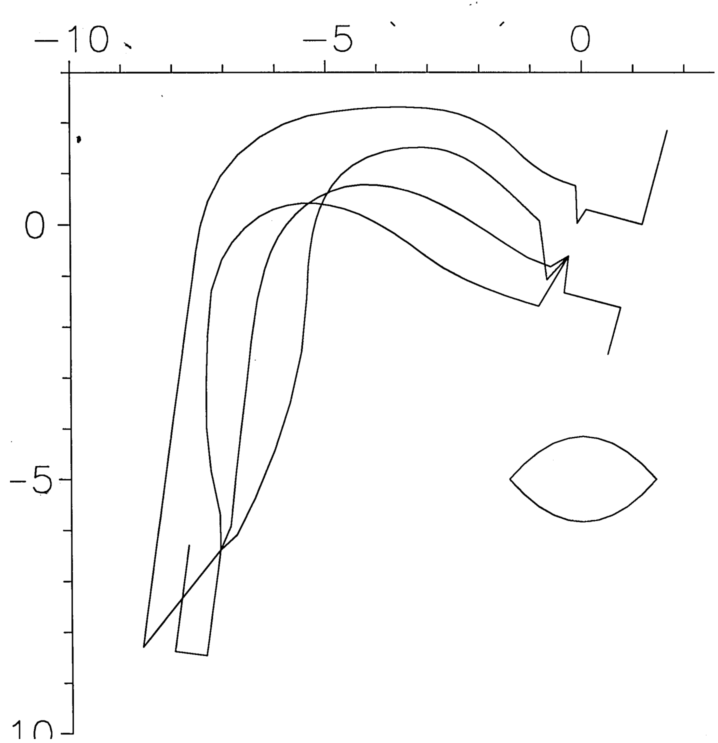




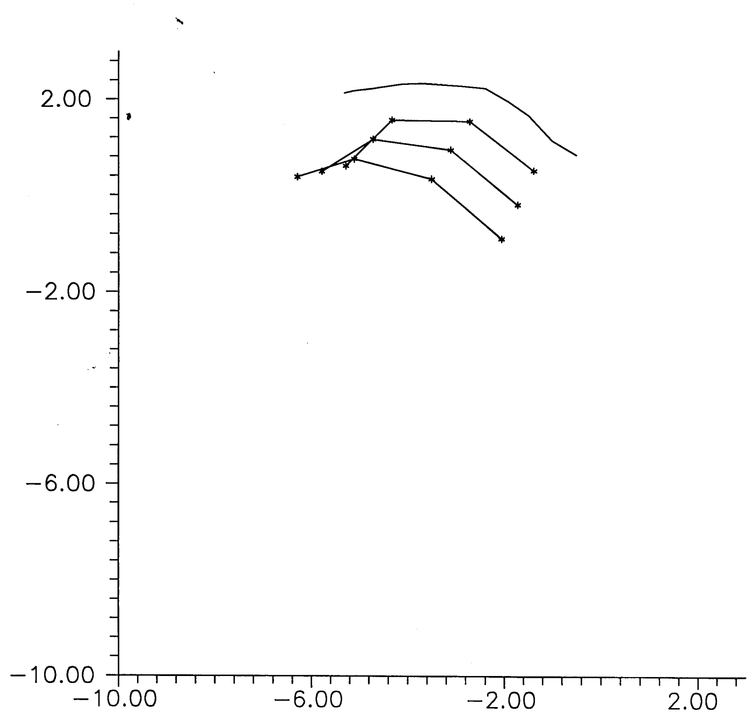
Relation of broken-line construction to space-filling circles



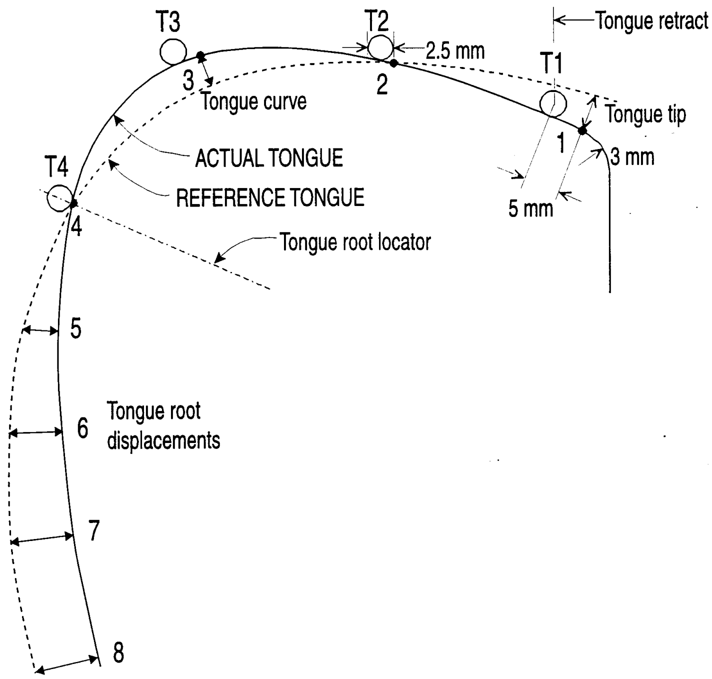
Articulation of ``front-raising'' basis function resulting from broken-line construction



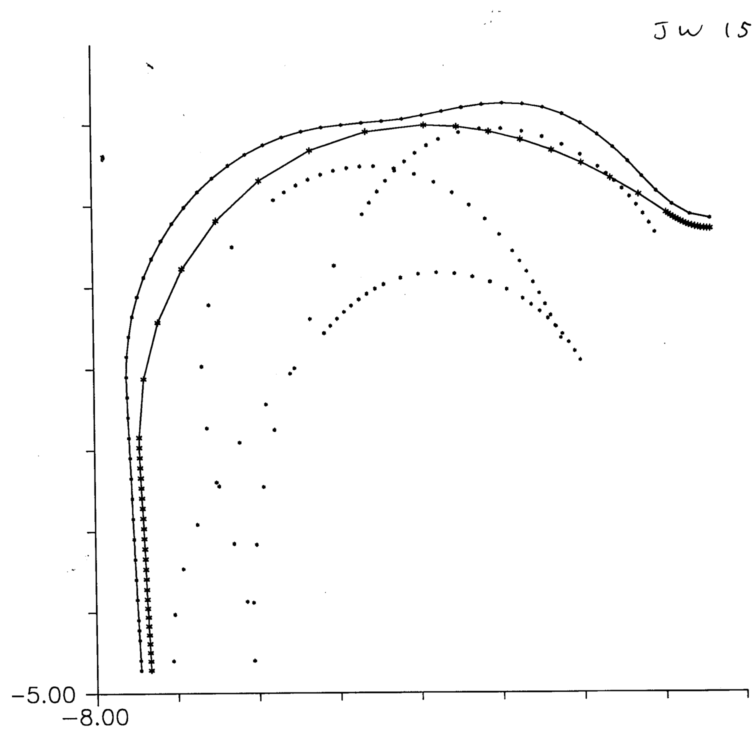
Articulation of /i/-/a/ seen in microbeam data



Refinement to 2-basis tongue outline to make shape corrections, articulate the tongue tip

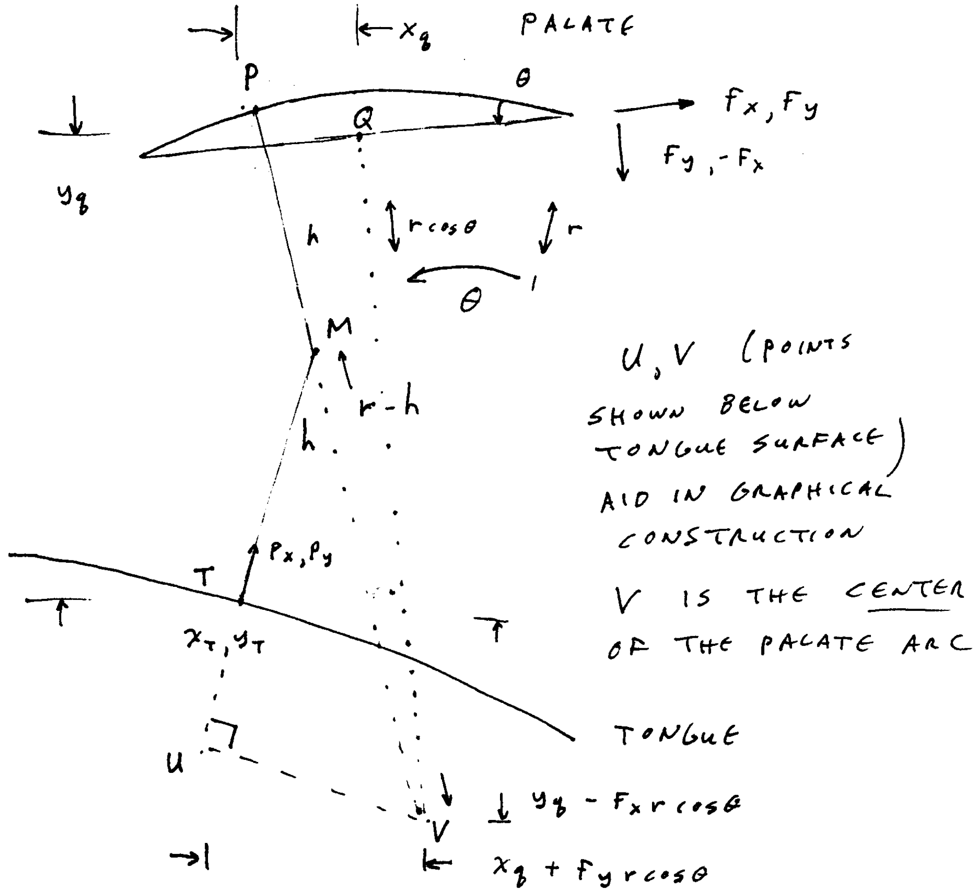


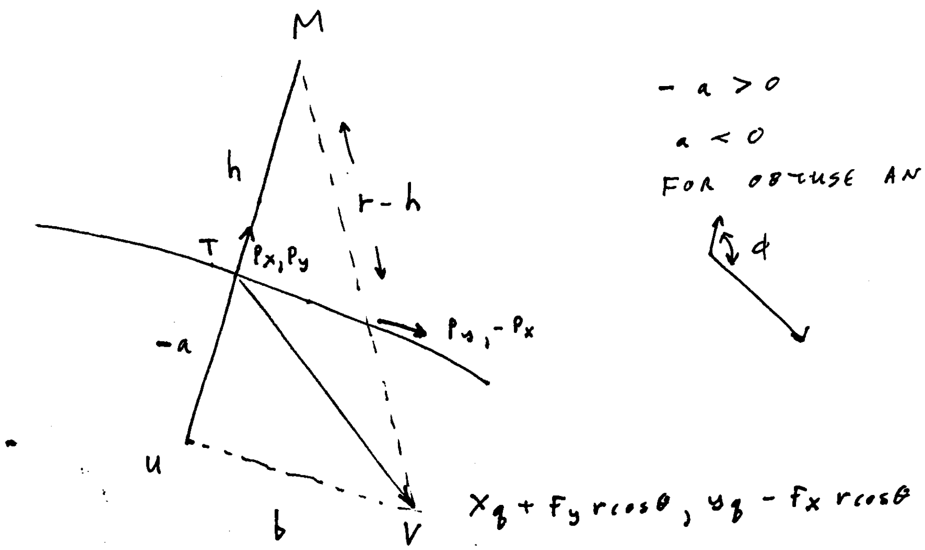
Introducing a smooth ``reference-palate'' for generating a 2-basis tongue outline with an adequate `` working space.''



Geometric construction of space-filling circle between palate and tongue outlines defined by piecewise circular arcs

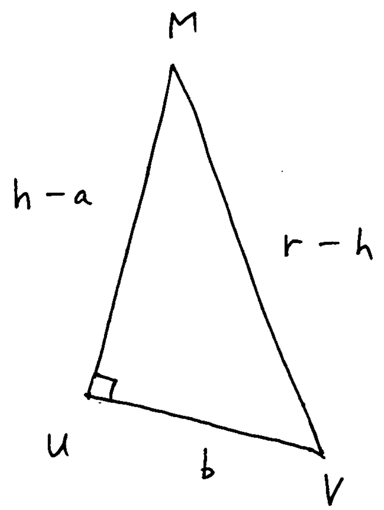
DETERMINING BROKEN-LINE DISTANCE h :





$$a = (x_q + F_y r\cos\theta)P_x + (y_q - F_x r\cos\theta)P_y$$

$$b = (x_q + F_y r\cos\theta)P_y - (y_q - F_x r\cos\theta)P_x$$



$$(h-a)^2 + b^2 = (r-h)^2$$

$$\cancel{h^2} - 2ha + a^2 + b^2 = r^2 - 2rh + \cancel{h^2}$$

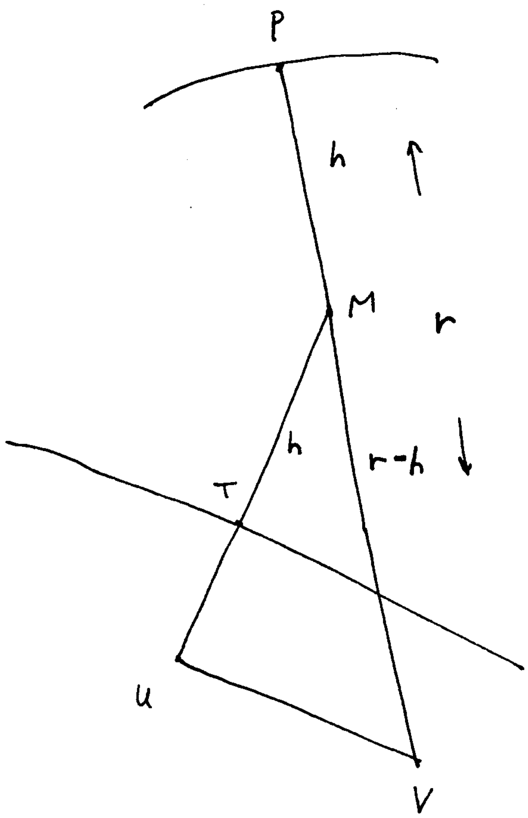
CANCELLATION AVOIDS QUADRATIC

$$h = \frac{a^2 + b^2 - r^2}{2(a-r)}$$

$$h = \frac{(F_x y_g - F_y x_g)\cos\theta - \left(x_g^2 + y_g^2 - \frac{d_m^2}{4}\right)\frac{1}{2r}}{1 + (F_x p_y - F_y p_x)\cos\theta - (p_x x_g + p_y y_g)\frac{1}{r}}$$

$$\frac{1}{r} = k_p \qquad \sin\theta = \frac{k_p d_m}{2}$$

V. MILENKOVIC AND P. MILENKOVIC,
TONGUE MODEL FOR CHARACTERIZING
VOCAL TRACT KINEMATICS (1996)
J LENARCIC AND V. PARENTI - CASTELLI
(EDS.) RECENT ADVANCES IN ROBOT
KINEMATICS, 217-224, KLUWER.



LOCATING POINT P:

$$P = M + \frac{\overleftarrow{MV}}{|\overleftarrow{MV}|} h$$

WHERE $|\overleftarrow{MV}| = r - h > 0$ FOR GEOMETRY SHOWN

$\{ r > 0 \quad r \leq h$ OUTSIDE "WORKING ENVELOPE" $|$

$r < 0$ TO BE CONSIDERED SEPARATELY

$$\text{SO} \quad P = M + \overleftarrow{MV} \frac{h}{r-h}$$

$$x_p = x_T + P_x h + \left(P_x h - (x_g + F_y r\cos\theta)\right)\frac{h}{r-h}$$

$$y_p = y_T + P_y h + \left(P_y h - (y_g - F_x r\cos\theta)\right)\frac{h}{r-h}$$

OR

$$x_p = x_T + \left(P_x r - F_y r\cos\theta - x_g\right)\frac{h}{r-h}$$

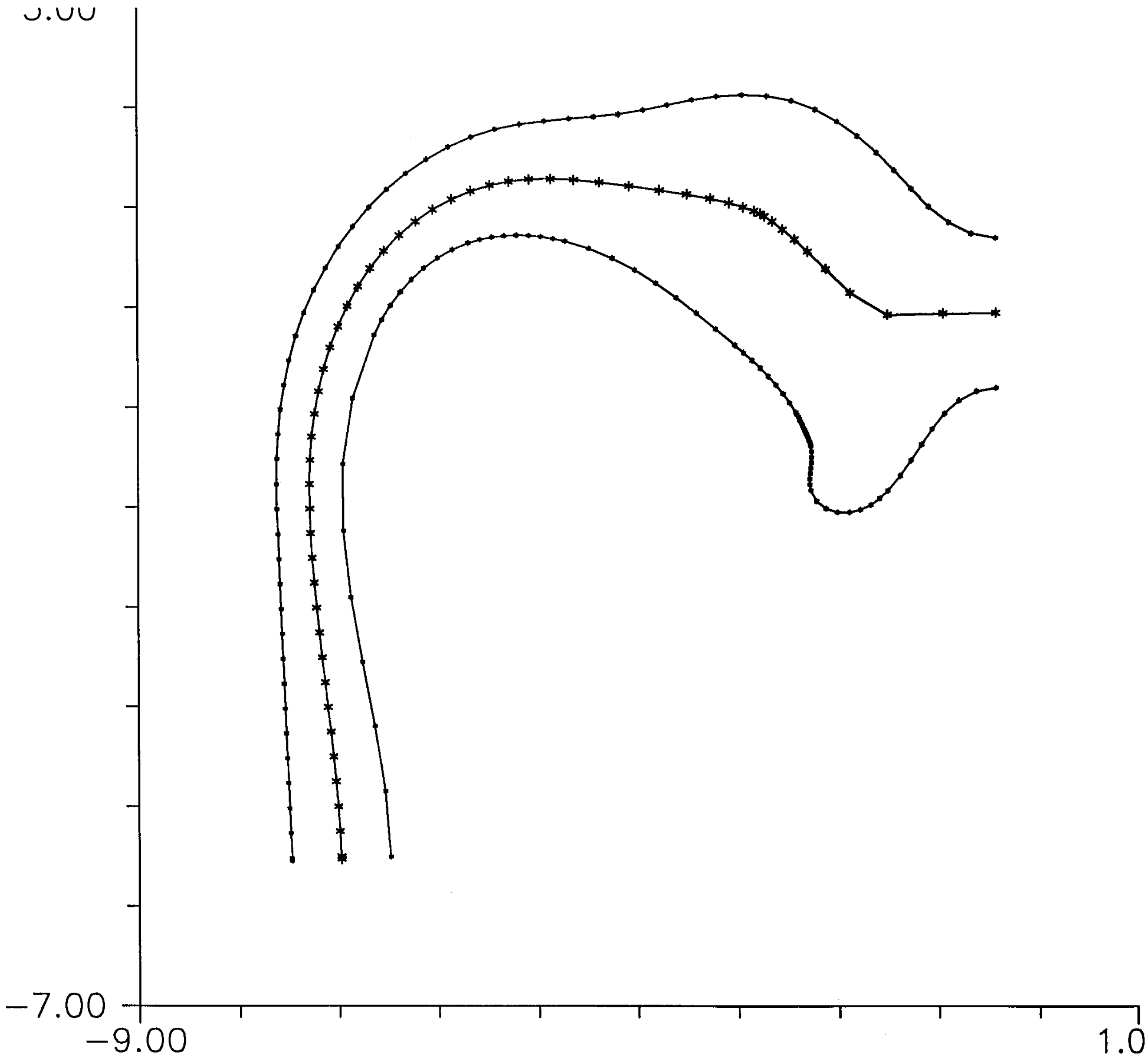
$$y_p = y_T + \left(P_y r + F_x r\cos\theta - y_g\right)\frac{h}{r h}$$

OR

$$x_p = x_T + \left(P_x - F_y\cos\theta - x_g/r\right)\frac{h}{1-\frac{h}{r}}$$

$$y_p = y_T + \left(P_y + F_x\cos\theta - y_g/r\right)\frac{h}{1-\frac{h}{r}}$$

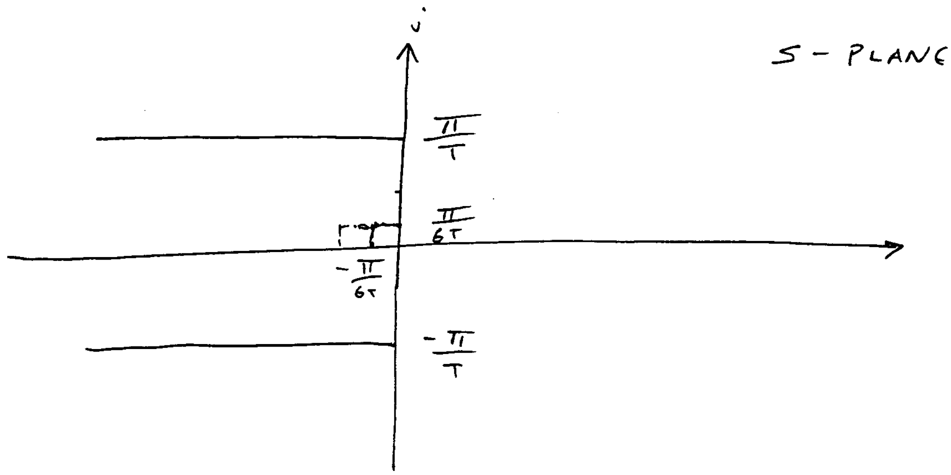
Vocal tract midline – centers of space-filling circles

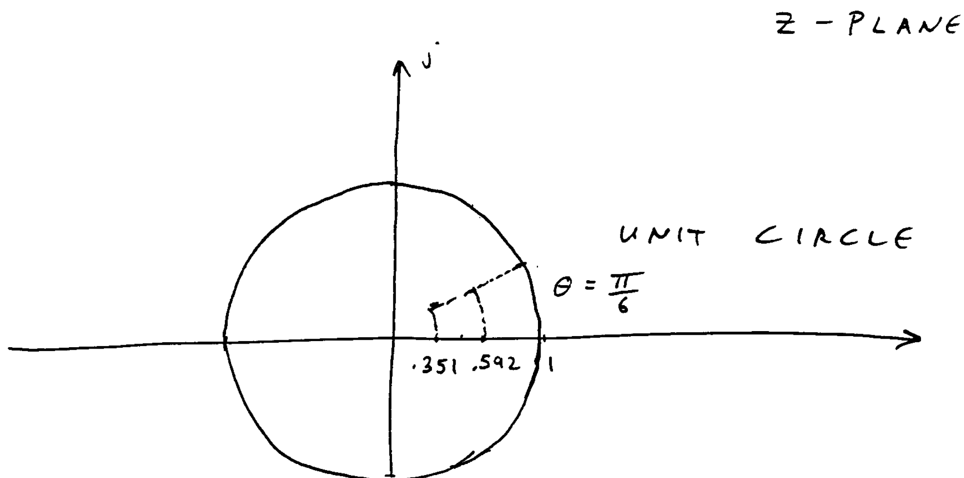


Acoustic streamlines follow a conformal map in the low-frequency approximation. Signal-processing engineers are familiar with this conformal map:

CONFORMAL TRANSFORMATION

$$z = e^{sT} \qquad s = \frac{1}{T} \ln z$$





$$z = r e^{j\theta}$$

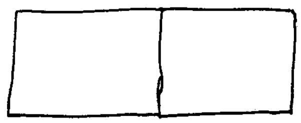
BACK-MAPPING:

$$s_R = \frac{1}{T} \ln r$$

$$s_I = \frac{1}{T} \theta$$

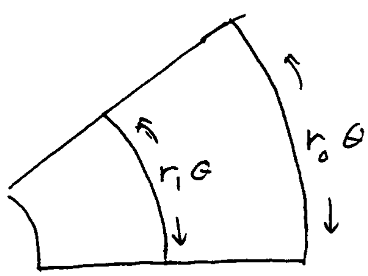
RADIAL SPACING OF FIELD LINES

$$S_R = \frac{1}{T} \ln r$$



S_{R2} S_{R1} S_{R0}

$$S_{R1} = \frac{1}{2}\left(S_{R2} + S_{R0}\right)$$



r_2 r_1 r_0

$$\frac{\ln r_1}{T} = \frac{1}{2}\left(\frac{\ln r_2}{T} + \frac{\ln r_0}{T}\right)$$

$$\ln r_1 = \ln r_2^{\frac{1}{2}} r_0^{\frac{1}{2}}$$

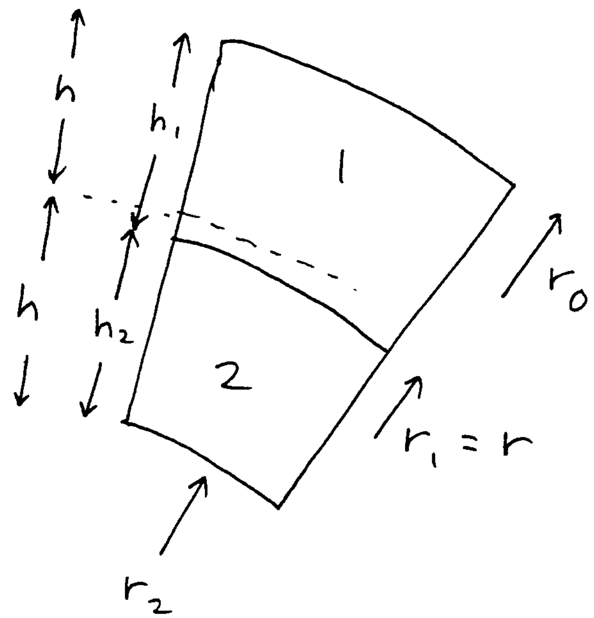
$$r_1 = \sqrt{r_2 \; r_0}$$

ALSO PRESERVES SCALE

SET $\dfrac{r_1 - r_2}{r_1\theta} = \dfrac{r_0 - r_1}{r_0\theta}$ $\therefore$ $\dfrac{r_2}{r_1} = \dfrac{r_1}{r_0}$

$\therefore$ $r_1 = \sqrt{r_2 r_0}$

EFFECT OF SCALE SIMILARITY

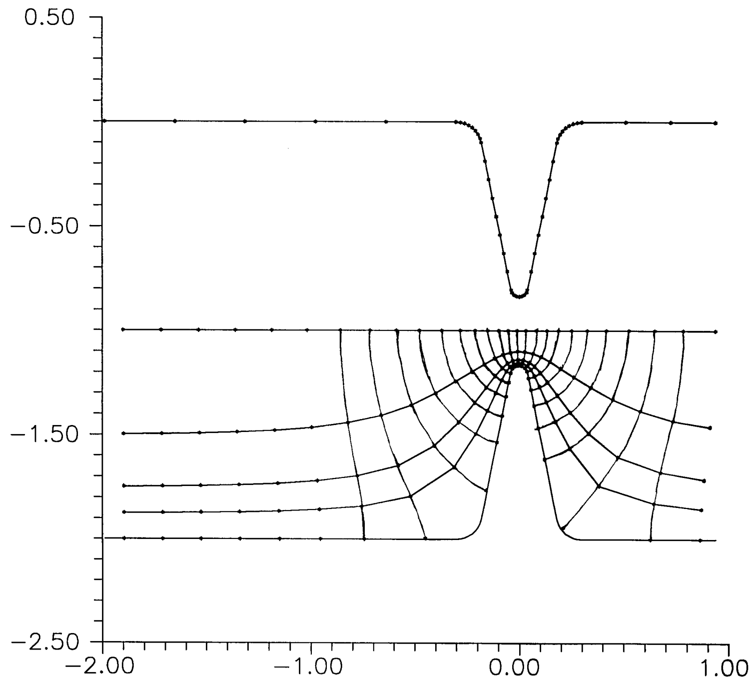


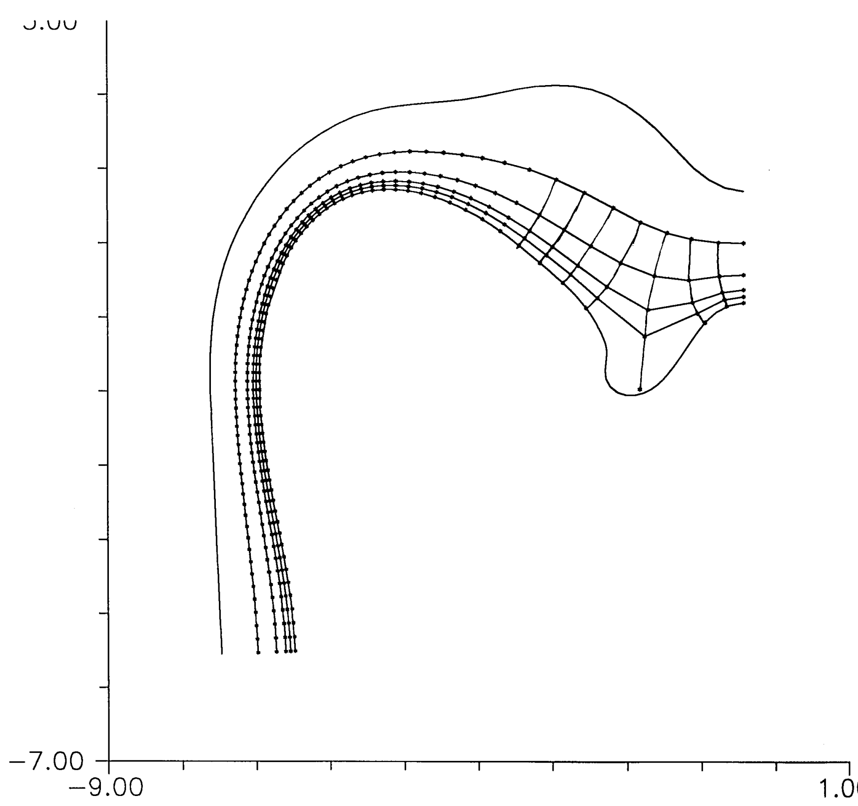
$$\frac{h_2}{r_2} = \frac{h_1}{r} \qquad \{ r = r_1 \}$$

$$\frac{h_2}{r - h_2} = \frac{h_1}{r}$$

$$\boxed{\frac{1}{h_2} = \frac{1}{r} + \frac{1}{h_1}}$$

Examples of graphical construction of approximation to the conformal map based on relation between local curvature and distance to boundaries for circular-geometry conformal map





Conclusion

These concepts are incorporated into the computer program XYCalc. Actual and reference palate outlines can be generated from Microbeam data using the program TF32. TF32 can also generate pellet position and formant frequency snapshots for use by XYCalc.