

---

# El Modelo Acústico

## 4.1. INTRODUCCIÓN

El modelo acústico representa la propagación de la onda sonora a través de los tractos vocal y nasal, hacia el ambiente exterior. A su vez, la onda sonora constituye una transmisión de perturbaciones en las partículas del aire, y en general, de cualquier medio suficientemente elástico. Las referidas perturbaciones consisten, concretamente, en variaciones locales de la presión, velocidad y desplazamiento de las partículas, respecto a los valores de equilibrio. En consecuencia, el modelo acústico debe reflejar la evolución temporal y espacial de los cambios en presión y velocidad de las partículas.

En general, existen tres tipos de modelos acústicos para síntesis articulatoria. El primero representa la propagación de la onda, hacia adelante y hacia atrás, en un tubo acústico ideal, sin pérdidas de energía. Este enfoque se denomina estructura de lattice o de Kelly-Lochbaum [31], y es el único que permite síntesis en tiempo real. También es el modelo con las emisiones sintéticas de más baja calidad, porque como se mencionó, omite las pérdidas en los tractos y otros efectos relativos a la interacción de éstos con las fuentes. Además, la función de longitud forzosamente debe mantenerse constante, lo cual empobrece la representación de varias configuraciones articulatorias.

El segundo enfoque propone un método híbrido en el dominio del tiempo y de la frecuencia [71]. En este caso, el filtro se modela en el dominio de la frecuencia, y la fuente, con sus características no-lineales, en el dominio del tiempo. No obstante, con estos modelos no resulta viable el cálculo de las presiones y velocidades, y también presenta dificultades con algunas transiciones entre configuraciones articulatorias.

El resto de las implementaciones corresponde a modelos en el dominio del tiempo, caracterizados por el empleo de sistemas de ecuaciones lineales que permiten el cálculo de la presión y la velocidad en diversos puntos de los tractos. Estos modelos generan las emisiones sintéticas de mayor calidad, incorporando las pérdidas de energía, y calculando las presiones y velocidades [16], aunque los cálculos resultan costosos, manifiestan distorsión de frecuencia [86], y revelan algunas carencias modelando componentes intrínsecamente dependientes de la frecuencia. La aptitud de este modelo para representar configuraciones articulatorias y las transiciones entre ellas [42, 71], constituye una gran ventaja para los propósitos de esta investigación, razón por la cual se adopta e implementa.

En la implementación de modelos en el dominio del tiempo, se distingue entre dos corrientes. La primera resuelve mediante métodos numéricos los sistemas de ecuaciones diferenciales que representan las variaciones de presión y velocidad del volumen a lo largo de los tractos [8]. La otra opción apela a una elegante analogía de cada cilindro en la aproximación de los tractos con una red eléctrica T [22, 42]. Esta última variante se adopta aquí, principalmente porque admite de forma directa el cómputo de las funciones de transferencia asociadas a funciones de área y longitud específicas.

Tales funciones de transferencia resultan imprescindibles para la caracterización de configuraciones articulatorias durante la inversión de las actividades de los músculos supraglotales.

En general, el resto del capítulo abarca la definición del modelo acústico, a partir de la integración de las descripciones y resultados presentados en la literatura de Acústica Teórica.

## 4.2. ECUACIÓN DE LA ONDA

La representación más directa y aceptada de las perturbaciones en las partículas del aire es la Ecuación de la Onda [10], que bajo algunos supuestos pertinentes se expresa como:

$$\begin{aligned} -\frac{\partial p}{\partial x} &= \frac{\rho}{A} \frac{\partial u}{\partial t} \\ -\frac{\partial u}{\partial x} &= \frac{A}{\rho c^2} \frac{\partial p}{\partial t} \end{aligned} \quad (4.1)$$

En la derivación de la Ecuación 4.1, se adoptan los siguientes supuestos [57]:

1. El medio es homogéneo, lo cual permite establecer  $\rho$  igual a la densidad atmosférica promedio.
2. El cambio de presión en pequeñas distancias puede linealizarse, y así, procede la supresión de todos los términos que no sean de primer orden en la expansión en serie de Taylor de la presión respecto al argumento  $x$ .
3. Las partículas del aire no experimentan viscosidad.
4. La velocidad de las partículas de aire es relativamente pequeña. De lo contrario, el término  $v \frac{\partial v}{\partial x}$  resultaría significativo respecto a  $\rho \frac{\partial v}{\partial t}$ , y la Ecuación 4.1 derivaría en no-lineal, difícil de resolver.
5. La producción sonora corresponde a un proceso adiabático.

En este sistema de ecuaciones diferenciales parciales lineales,  $A$  es la función de área de la sección transversal del tubo o conducto de propagación, y las variables  $p$  y  $u$  representan las variaciones en presión y velocidad del volumen de las partículas. Además, la Ecuación 4.1 supone una propagación unidimensional de la onda, sin cambios en las dimensiones  $y$  o  $z$ . Por su parte, la velocidad del volumen es la tasa de flujo de partículas perpendicular al área  $A$ .

A partir de la Ecuación 4.1 se establece una analogía con la propagación de una onda eléctrica en una línea de transmisión, donde para un voltaje  $e(x, t)$  y corriente  $i(x, t)$ , se verifican las siguientes relaciones [21, 22]:

$$\begin{aligned} -\frac{\partial e}{\partial x} &= L \frac{\partial i}{\partial t} \\ -\frac{\partial i}{\partial x} &= C \frac{\partial e}{\partial t} \end{aligned} \quad (4.2)$$

$L$  y  $C$  constituyen la inductancia y capacitancia por unidad de longitud de la línea de transmisión. En este sentido,  $\rho/A$  y  $A/\rho c^2$  se denominan *inductancia acústica* y *capacitancia acústica*, respectivamente. Nótese, sin embargo, que en la Ecuación 4.1 se asume que el tubo acústico tiene una sección transversal de dimensiones constantes, lo cual, aunque simplifica la expresión, proporciona

una aproximación muy pobre para los sonidos del habla. Empero, una solución razonable puede obtenerse conectando varios cilindros de área y longitud variable, para aproximar un tracto vocal de área no uniforme, correspondiente a la discretización de  $A(x, t)$  discutida en el capítulo precedente. Consecuentemente, sobre la base de la analogía eléctrica, el tracto vocal se modelará mediante una serie de redes T conectadas en cascada. La Figura 4.1 ilustra estos comentarios. Esta discretización puede afectar la ubicación de resonancias hasta en un 8% [70], factor de escasa importancia.

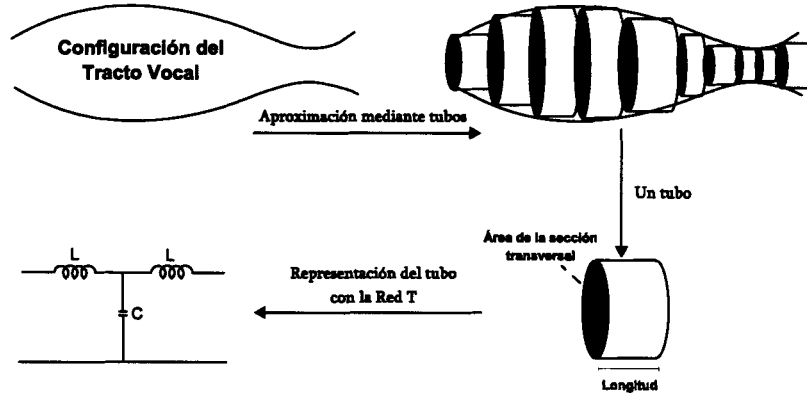


Figura 4.1: Aproximación del Tracto Vocal.

### 4.3. PÉRDIDAS DE ENERGÍA

En [56], Portnoff introdujo algunos cambios en la Ecuación 4.1 con el fin de considerar las pérdidas de energía resultantes de la vibración de las paredes de los tractos, que constituyen tejidos blandos, elásticos, susceptibles de movimiento por efecto de la presión sonora. Evidentemente, la noción de área constante resulta insostenible, si bien puede asumirse que  $A(x, t)$  cambia *lentamente* en espacio y tiempo. Con esta última condición, la propagación del sonido en un tubo no uniforme se expresa:

$$\begin{aligned} -\frac{\partial p}{\partial x} &= \rho \frac{\partial u}{\partial t} \\ -\frac{\partial u}{\partial x} &= \frac{1}{\rho c^2} \frac{\partial p}{\partial t} + \frac{\partial A}{\partial t} \end{aligned} \quad (4.3)$$

Este sistema carece de solución en forma cerrada excepto para las configuraciones más simples [58]. Las pérdidas introducidas por la vibración de las paredes se modelan mediante la Ecuación 4.4, acoplada a la Ecuación 4.3.

$$m_w \frac{d^2}{dt^2} (\delta A(x, t)) + b_w \frac{d}{dt} (\delta A(x, t)) + k_w \delta A(x, t) = p(x, t) \quad (4.4)$$

Se considera que toda zona de tejido se encuentra funcionalmente separada de las demás (la vibración no es distribuida), y que los coeficientes permanecen constantes y uniformes a lo largo de los tractos. Los valores a utilizar son:  $m_w = 1.5$  gm,  $b_w = 1400$  gm/s y  $k_w = 30000$  dinas/cm [16,42]. En la Ecuación 4.4,  $\delta A(x, t)$  representa la variación del área respecto a su valor nominal  $A_0(x, t)$ :

$$A(x, t) = A_0(x, t) + \delta A(x, t) = A_0(x, t) + y(x, t)S_0(x, t) \quad (4.5)$$

$S_0(x, t)$  es el perímetro de la sección tubular considerada, y  $y(x, t)$  es el desplazamiento de la pared, en el sentido de la normal, como resultado de la presión [57].

El efecto de las pérdidas viscosas y térmicas también debe considerarse. Aunque tales pérdidas resultan de magnitud inferior a las producidas por vibración de las paredes, su efecto adquiere importancia en las frecuencias altas, donde se manifiestan más la fricción y el calor. Estas pérdidas se modelan introduciendo en la red T las resistencias  $R$  y  $G$ , para la fricción viscosa y para la disipación térmica, respectivamente [16, 57].

#### 4.4. CONDICIONES DE FRONTERA

Por cuanto la impedancia mecánica en la glotis resulta mucho mayor que la del sistema acústico supraglotal, la fuente puede separarse, en principio, del resto del sistema [28, 42]. En concordancia con la teoría Fuente-Filtro de producción de la voz [21], el modelo acústico se constituye en un filtro de la energía proveniente del sistema glotal. La primera frontera del modelo se ubica en la glotis, justo al inicio del tracto faríngeo:

$$P_g = p(x_0, t) \quad (4.6)$$

donde  $P_g$  es la presión glotal, y  $x_0$  la coordenada en la frontera. En su tránsito al exterior, si el paso velofaríngeo se encuentra abierto<sup>1</sup>, la energía sonora se distribuye entre el tracto vocal y el nasal en el punto  $x_K$ , donde ocurre la bifurcación nasal. Allí, las relaciones de velocidad y presión son las siguientes [42]:

$$\begin{aligned} u(x_K^-, t) &= u(x_K^+, t) + u'(0, t) \\ p(x_K^-, t) &= p(x_K^+, t) = p'(0, t) \end{aligned} \quad (4.7)$$

$p'$  y  $u'$  denotan la presión y velocidad del volumen en el tracto nasal. En la bifurcación nasal culmina el tracto faríngeo e inicia el tracto oral, hasta los labios en la longitud  $x = x_M$ . A su vez, el tracto nasal inicia en  $x_K$  y finaliza en las fosas nasales,  $x_N$ . En lo sucesivo, las funciones de área y longitud nasal se mantienen relativamente invariables, con 11 cilindros aproximando un tracto nasal de 11.0 cm de longitud. La Figura 4.2 exhibe la distribución del área nasal.

Finalmente, el sonido es radiado al exterior en  $x_M$  o  $x_N$ . Para modelar este proceso, se recurre a la aproximación por circuito paralelo de Flanagan [22], cuya ventaja fundamental respecto a otros modelos es que las ecuaciones no incluyen términos dependientes de la frecuencia, lo que posibilita la simulación en el dominio del tiempo. Así, la radiación viene modelada por:

$$u(x_M, t) = \int_0^t S_{rad}(\lambda) p(x_M, \lambda) d\lambda + G_{rad}(t) p(x_M, t) \quad (4.8)$$

donde  $S_{rad}(t) = 9\pi A_0(x_M, t)/128\rho c^2$  y  $G_{rad}(t) = 3\pi\sqrt{\pi A_0(x_M, t)}/8\rho$ . En definitiva, la Figura 4.3 resume la aproximación del tracto vocal, con fuente  $U_g$ , y el modelo de radiación de Flanagan.

La expresión de la radiación nasal adopta la forma de la Ecuación 4.8, con  $x_M$  reemplazado por  $x_N$ .

<sup>1</sup>Ocurre durante el análisis y síntesis de los fonemas /m/ y /n/, donde se igualará  $|E_5 E_6|$  a 0.5 cm, y el área del paso velofaríngeo al valor promedio de 0.5 cm<sup>2</sup> [78].

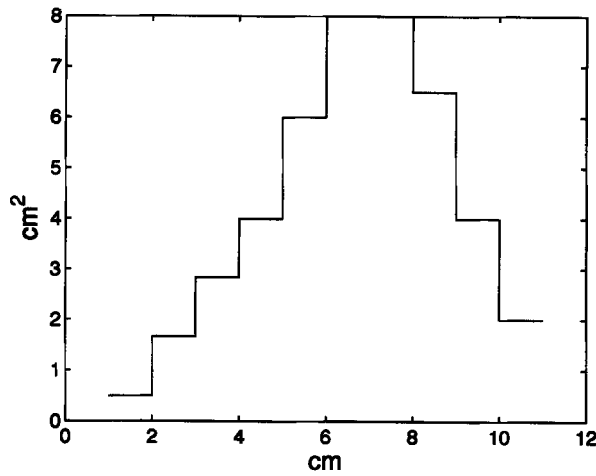


Figura 4.2: Función de Área nasal [43].

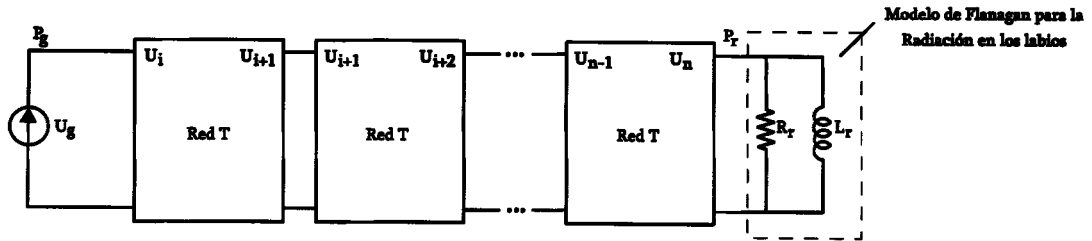


Figura 4.3: El Tracto Vocal como una serie de Redes T.

### 4.5. DISCRETIZACIÓN DEL MODELO

Según la teoría revisada, cada cilindro en la aproximación de los tractos, incluyendo las pérdidas de energía, equivale a la red T de la Figura 4.4.

Las expresiones para cada componente del circuito son [16]:

$$R(\omega) = \frac{S\sqrt{v\mu\omega}}{2\sqrt{2}A^2}l$$

$$L = \frac{v}{2A}l$$

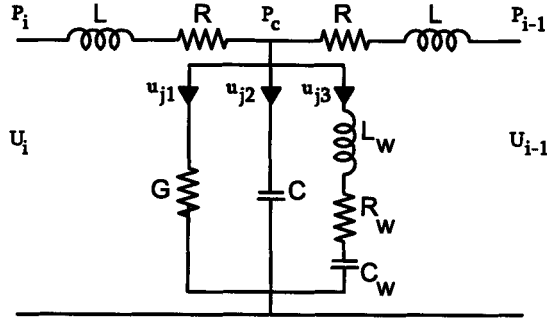
$$C = \frac{A}{vc^2}l$$

$$G(\omega) = \frac{(\eta - 1)S}{vc^2} \sqrt{\frac{\lambda\omega}{2\xi v}}l$$

$$R_w = \frac{b}{S^2l}$$

$$L_w = \frac{m}{S^2l}$$

$$C_w = \frac{S^2l}{k}$$



**Figura 4.4:** Representación de un cilindro mediante una Red T, después de introducir las pérdidas de energía.

donde

$S$  :  $S = 2\sqrt{A\pi}$  es la circunferencia del cilindro.

$A$  y  $l$  : área y longitud del cilindro, respectivamente.

$\mu$  : viscosidad, igual a  $1.86 \times 10^{-4}$  dina-s/cm<sup>2</sup>.

$\lambda$  : coeficiente de conducción calórica del aire, igual a  $0.55 \times 10^{-3}$  cal/cm-s-grad.

$\eta$  : constante de gas adiabático, 1.4.

$\xi$  : calor específico, 0.24 cal/gm-grad.

$\omega$  : frecuencia angular (rad/s).

Como se aprecia, las resistencias  $R(\omega)$  y  $G(\omega)$  dependen de la frecuencia angular. Para la simulación en el dominio del tiempo,  $\omega$  se fija en 2.5 kHz, constante que afecta levemente la ubicación y ancho de banda de las resonancias del tubo acústico [16].

A partir de la Figura 4.4 se deriva directamente el sistema de ecuaciones 4.9, relacionando las presiones y velocidades del volumen de una red  $j$  con la de redes vecinas. Para la red  $j$ ,  $P_j(t)$  equivale a  $P_c(t)$  en el centro de la red, mientras que la corriente  $U_j(t)$  se encuentra a la entrada del circuito. Recuérdense las analogías entre la presión de aire y el voltaje, y entre la velocidad del volumen y la corriente. Además,  $u_{j3}$  identifica a la corriente en los componentes eléctricos que modelan las pérdidas derivadas de la vibración de las paredes de los tractos.

$$\begin{aligned}
 P_{j-1}(t) - P_j(t) &= [L_{j-1}(t) + L_j(t)] \frac{d}{dt} U_j(t) + [R_{j-1}(t) + R_j(t)] U_j(t) \\
 U_j(t) - U_{j+1}(t) &= u_{j1}(t) + u_{j2}(t) + u_{j3}(t) \\
 &= G_j(t) P_j(t) + C_j(t) \frac{d}{dt} P_j(t) + u_{j3}(t) \\
 P_j(t) &= R_{w,j}(t) u_{j3}(t) + L_{w,j}(t) \frac{d}{dt} u_{j3}(t) + \int_0^t \frac{u_{j3}(\tau)}{C_{w,j}(\tau)} d\tau
 \end{aligned} \tag{4.9}$$

Los términos anteriores son funciones continuas que deben discretizarse para implementar las ecuaciones computacionalmente. En la Ecuación 4.9 hay tres tipos diferentes de términos: simple, diferencial e integral, generalizados en la Ecuación 4.10.

$$\begin{aligned}
y_1(t) &= c_1(t)x(t) \\
y_2(t) &= \frac{d}{dt}c_2(t)x(t) \\
y_3(t) &= \int_0^t c_3(t)x(t)dt
\end{aligned} \tag{4.10}$$

La ecuación simple se discretiza directamente. Y Maeda [42] sugiere la aplicación de la regla trapezoidal para aproximar  $y_i(t)$  en las otras dos ecuaciones. Luego,  $t$  se evaluará en intervalos determinados por el período de muestreo constante  $T_s$ . Naturalmente, la frecuencia de muestreo equivale a  $1/T_s$ , y se establecerá en 20 kHz, valor que proporciona un equilibrio entre la distorsión de frecuencia introducida por la discretización y la cantidad de tiempo requerida para resolver las ecuaciones anteriores, en todos los cilindros [42].

Recordando que debe muestrearse en  $t = nT_s$  ( $n = 1, 2, 3, \dots$ ), las ecuaciones diferenciales anteriores, para todos los cilindros, se transforman y agrupan en tres sistemas de ecuaciones algebraicas lineales, modelando la dinámica de los tractos faríngeo, oral, y nasal. Posteriormente, mediante eliminación gaussiana, se calculan  $U_j(nT_s)$  y  $P_j(nT_s)$ , para  $1 \leq j \leq 60$  en el tracto vocal, y  $1 \leq j \leq 11$  en el tracto nasal. La salida del sistema, o señal verbal sintetizada, viene dada por la diferencia entre las velocidades del volumen en los puertos de radiación en un tiempo dado, y el valor en el instante o muestra previa [22].

## 4.6. FUENTES DE ENERGÍA

En la síntesis de fonemas sonoros,  $U_g$  se equipara a la salida de las redes del Capítulo 2, distinguiendo entre el locutor masculino y el femenino. No obstante, la síntesis de fricativas exige la representación del ruido generado en los tractos supraglotales. En este sentido, la sección siguiente describe un modelo funcional, que añade la turbulencia justo en la sección del tracto donde se produce. Por su parte, la fuente de las oclusivas /p/ y /b/, sintetizadas con fines de evaluación (Capítulo 5), se simula cerrando el tracto oral por un breve instante de tiempo, de tal manera que los componentes de la red acumulen energía, y la liberen cuando la oclusión cese.

### 4.6.1. FUENTE DE TURBULENCIA

Cuando el aire fluye a través de una constricción, surgen fluctuaciones aleatorias de velocidad en el flujo que pueden actuar como una fuente de energía denominada *turbulencia*. Específicamente, para las consonantes /f/ y /s/, la energía proviene de la colisión del volumen de aire con un obstáculo más anterior. Por ejemplo, en la /s/ se forma una constricción alveolar, pero la energía se introduce, principalmente, por la colisión del volumen con los incisivos [78].

La intensidad espectral de este ruido viene dada por el número de Reynolds  $R_e$  [41]. Si  $R_e > R_{ec}$ , entonces se genera la turbulencia.  $R_{ec}$  es el número crítico de Reynolds, fijado durante la síntesis en 2700 [16], mientras que  $R_e$  se define como

$$R_e = \frac{\rho v h}{\mu} \tag{4.11}$$

$v$  es la velocidad lineal del flujo de aire ( $1000 \text{ cm}^3/\text{s}$  [16]),  $h$  es la dimensión característica (área) de la constricción, y  $\mu$  es la viscosidad del aire [72]. Concretamente, la turbulencia  $U_{jt}$  creada en una sección  $j$ , se modelará a través de la relación  $U_{jt} = P_{jt}/R_{jt}$  [67].  $P_{jt}$  viene dada por

$$\begin{aligned}
 P_{jt} &= (\text{turbg})(\text{rand})(R_e^2 - R_{ec}^2) \text{ si } R_e > R_{ec} \\
 &= 0 \text{ si } R_e \leq R_{ec}
 \end{aligned}
 \tag{4.12}$$

donde *turbg* se establece en  $20 \times 10^{-6}$  como ganancia de la turbulencia [16], y *rand* es un número aleatorio uniformemente distribuido entre  $-0.5$  y  $0.5$ . A su vez, la resistencia  $R_{jt}$  de la fuente es

$$R_{jt} = \frac{\rho |\bar{U}|}{2A^2} \tag{4.13}$$

donde  $\bar{U}$  constituye una versión filtrada de la velocidad del volumen  $U$ :

$$\bar{U}(n) = \bar{U}(n-1) + [U_j(n) - \bar{U}(n-1)]2\pi f_g T_s \tag{4.14}$$

La frecuencia  $f_g$  no es crítica: se establecerá en 2 kHz [16]. Finalmente, para efectos de la síntesis,  $U_{jt}$  se acopla en paralelo a la entrada de la sección  $j$  donde ocurre la colisión, con el propósito de recrear la turbulencia.



---

# Aprendizaje de la Actividad Muscular

## 5.1. INTRODUCCIÓN

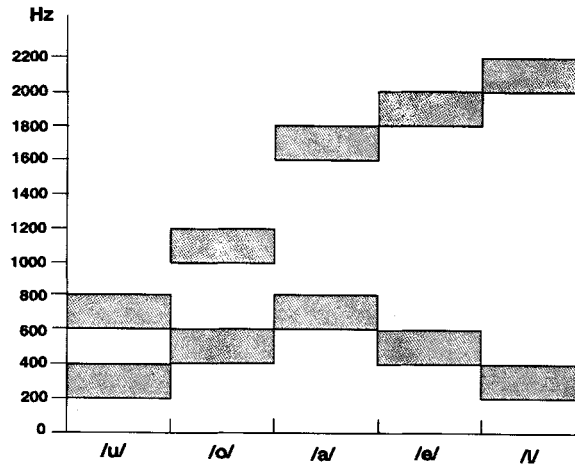
La inversión articulatoria, como se discute en el Capítulo 3, plantea el aprendizaje de la actividad del grupo de músculos supraglotales en  $p(t)$ , con el propósito de minimizar la distancia acústica entre una señal objeto  $S_v$  y la emisión sintética  $\phi(p(t))$ . En consecuencia, el aprendizaje equivale a la minimización de la Ecuación 1.1. El resto del capítulo formaliza la noción de distancia acústica y describe los métodos para el aprendizaje de la actividad muscular vinculada a los fonemas de interés. Para la inversión se aplican, específicamente, los Algoritmos Genéticos Continuos [24]. La definición del algoritmo en la sección 5.4, incluyendo estructura del cromosoma, operadores, funciones objetivo y criterio de convergencia constituyen propuestas originales. Naturalmente, la formación del corpus de señales objeto a invertir, los experimentos desarrollados, y las evaluaciones objetivas y subjetivas, abordadas en las secciones 5.5 y 5.6, también son propias.

La definición de distancia acústica entre dos señales dadas depende del tipo de fonema abordado. Para las vocales, las resonancias del tracto vocal aparecen particularmente nítidas en el dominio de la frecuencia, y reciben el nombre de *frecuencias formantes*. Como los formantes reflejan la configuración del tracto, los cambios en la función de área ocasionan variación en la estructura formántica. En este sentido, la Figura 5.1 aproxima las dos primeras resonancias de cada vocal española. En la práctica, esos valores difieren entre locutores y dependen del contexto de la vocal. Las resonancias en la Figura 5.1 distinguen las vocales en el dominio de la frecuencia, aunque las resonancias superiores también aportan información útil para la caracterización de las vocales, e introducen las peculiaridades acústicas del locutor.

En otros fonemas sonoros, como las consonantes nasales, la composición espectral incluye antiresonancias que dificultan la ubicación de las resonancias de la señal. En la /m/ y la /n/, las resonancias dependen del tracto faríngeo y del nasal, mientras que las antiresonancias proceden de una caída de energía en las frecuencias donde la impedancia del tracto delimitado por el velo y la oclusión oral resulta igual a cero. Tal composición espectral complica la caracterización de estos fonemas.

En relación con las consonantes fricativas /f/ y /s/, la caracterización adquiere otra forma. En la articulación de estos sonidos se crea una constricción que altera el flujo de aire, creando una fuente de turbulencia. Por tal razón, en el dominio de la frecuencia las fricativas muestran una composición de ruido coloreado. Para la /f/, la energía se distribuye de modo relativamente regular, mientras que para la /s/ suele concentrarse en las regiones superiores, aproximadamente alrededor de los 5 kHz [78, 89].

Las propiedades acústicas de los fonemas no abordados en esta investigación también resultan muy particulares. Por ejemplo, el espectrograma de las oclusivas (seguidas por una vocal) reproduce los siguientes eventos acústicos [21]:



**Figura 5.1:** Estructura formántica aproximada de las vocales españolas [50].

1. Tensión.
2. Breve impulso inicial en la velocidad del volumen.
3. Turbulencia en la constricción a medida que las paredes del tracto vocal retornan a su posición de equilibrio.
4. Turbulencia en la glotis por breve tiempo (aspiración).
5. Transición hacia la vocal (sonorización).

Obviamente, tal serie de eventos acústicos resulta imposible de invertir con enfoques estáticos como el propuesto en el presente estudio.

En concordancia con lo expuesto, y retornando a los fonemas de interés, la inversión depende del tipo de señal objeto, porque las características de vocales, nasales y fricativas difieren significativamente. En el caso de las vocales y nasales, dichas características pueden derivarse de las configuraciones articulatorias sobre el plano medial a través del análisis directo de la función de transferencia, tema de la próxima sección.

## 5.2. CÓMPUTO DE LA FUNCIÓN DE TRANSFERENCIA

Las resonancias y antiresonancias de los tractos corresponden respectivamente, y con un error despreciable, a los polos y ceros de la función de transferencia  $H(\omega) = U_0/U_g$  [41].  $U_g$  es la señal de excitación glotal, y  $U_0$  representa la señal radiada al EAL en los labios (vocales) o en las fosas nasales (consonantes nasales). A partir de las redes T puede obtenerse una expresión en forma cerrada de la función de transferencia, lo cual constituye una ventaja decisiva de este tipo de modelado. La derivación de  $H(\omega)$  inicia con el conjunto de Ecuaciones 5.1 y 5.2, que expresan las impedancias de la Figura 5.2.  $l_i$  es la evaluación de la función de longitud en la sección  $i$ .

$$\begin{aligned} Z_1 &= Z_i \tanh \Gamma_i/2 \\ Z_2 &= \frac{Z_i}{\sinh \Gamma_i} \end{aligned} \quad (5.1)$$

$$\begin{aligned}\Gamma_i &= \sqrt{(R + j\omega L)(G + j\omega C + 1/Z_w)} \cdot l_i \\ Z_i &= \sqrt{\frac{R + j\omega L}{G + j\omega C + 1/Z_w}} \\ Z_w &= R_w + j\omega L_w + \frac{1}{j\omega C_w}\end{aligned}\quad (5.2)$$

Luego, la Ecuación 5.3 establece la relación entre el voltaje y la corriente a la entrada y salida del circuito en la Figura 5.2.

$$\begin{bmatrix} P_i \\ U_i \end{bmatrix} = M_i(\omega) \begin{bmatrix} P_{i-1} \\ U_{i-1} \end{bmatrix} = \begin{bmatrix} \cosh \Gamma_i & Z_i \sinh \Gamma_i \\ Z_i^{-1} \sinh \Gamma_i & \cosh \Gamma_i \end{bmatrix} \begin{bmatrix} P_{i-1} \\ U_{i-1} \end{bmatrix}\quad (5.3)$$

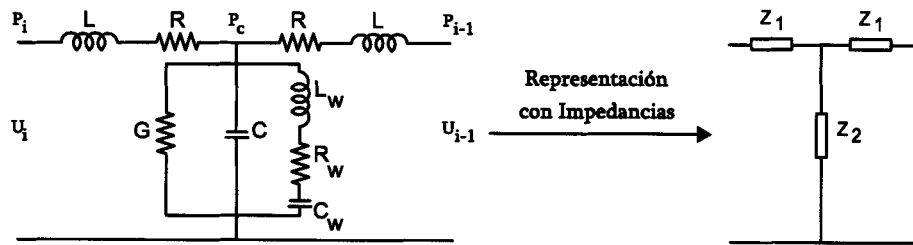


Figura 5.2: Representación con impedancias de la Red T.

De esta forma, toda sección o cilindro  $i$  en los tractos tiene asociada una matriz  $M_i(\omega)$ , denominada *matriz ABCD*. El cálculo de  $H(\omega)$  exige la multiplicación de las matrices ABCD de todas las secciones que representan los tractos. Numéricamente, la función de transferencia debe evaluarse para diversos valores de  $\omega$ , por lo que resulta más eficiente calcular la función de transferencia local,  $U_{i-1}/U_i$ , formando el producto de las funciones locales para obtener la función global de transferencia evaluada en un  $\omega$  dado. Así, se suprime el requisito de la multiplicación matricial, y la función de transferencia viene dada por la Ecuación 5.4 [2].

$$H(\omega) = \frac{U_0}{U_g} = \frac{U_{g-1}}{U_g} \dots \frac{U_{i-1}}{U_i} \dots \frac{U_0}{U_1}\quad (5.4)$$

El cálculo de los factores  $U_{i-1}/U_i$  requiere la impedancia del tubo acústico al frente de la sección  $i$ , dada por

$$Z_{T,i-1} = \frac{P_{i-1}}{U_{i-1}} = \frac{Z_{i-1}}{\sinh \Gamma_{i-1}} \left[ \cosh \Gamma_{i-1} - \frac{U_{i-2}}{U_{i-1}} \right]\quad (5.5)$$

Los fonemas /m/ y /n/, sin embargo, ameritan un tratamiento especial. Primero, en la producción de estas señales ocurre una oclusión en el tracto oral<sup>1</sup> cuyo modelado con un circuito abierto se ilustra en la Figura 5.3. Aquí, la radiación se expresa  $Z_r = Z_{OC} \coth \Gamma_{OC}$ , donde  $OC$  es el índice de la sección cerrada.

<sup>1</sup>En los labios y en los alvéolos para /m/ y /n/, respectivamente.

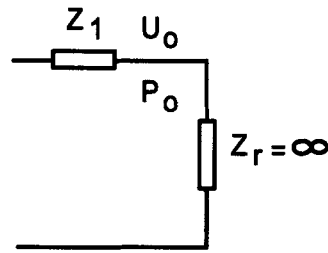


Figura 5.3: Modelo de oclusión labial.

Para las vocales sólo interesa el tracto vocal. Además, como se trata de configuraciones relativamente abiertas, la radiación recurre al modelo de Flanagan. En este caso, los factores  $U_{i-1}/U_i$  no incluyen ceros, pero sí polos, y en consecuencia,  $H(\omega)$  sólo contiene polos. Y dichos polos señalan las frecuencias formantes. De este modo, para caracterizar cualquier vocal sobre el plano medial basta computar la Ecuación 5.4, lo cual exime del alto costo computacional de producir la señal sintética y luego derivar sus picos espectrales con alguna técnica de rastreo de formantes. Por su parte, como en las nasales el paso velofaríngeo se encuentra abierto, la mayor parte de la energía acústica transita desde el tracto faríngeo hacia el nasal. No obstante, el acoplamiento con el tracto oral introduce ceros en la función de transferencia, en las frecuencias donde la impedancia del tracto oral también es nula. El tracto vocal acoplado se representa con una impedancia shunt  $Z_B$  insertada justo en la sección de bifurcación; el valor de  $Z_B$  se calcula a partir de la Ecuación 5.5. La Figura 5.4 ilustra la situación estudiada con la bifurcación ocurriendo en la sección  $i$ .  $Z_B$  es la impedancia del tracto oral, y  $Z_N$  la del nasal. Por simple inspección:

$$\frac{U_i}{U_{s,i}} = \frac{Z_B}{Z_B + Z_N} \tag{5.6}$$

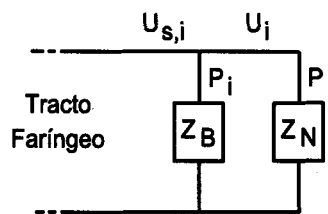


Figura 5.4: Impedancia shunt como modelo del tracto oral cerrado.

Finalmente, para calcular la función de transferencia de las nasales debe agregarse  $U_i/U_{s,i}$  y  $U_{s,i}/U_{i+1}$  a los factores de la Ecuación 5.4. Nótese que ahora  $H(\omega)$ , además de polos, incluye los ceros de  $Z_B$ . Estos ceros representan antiresonancias que como su nombre sugiere, pueden debilitar o cancelar las resonancias, razón por la cual el rastreo de formantes en la señal objeto tiene algunas dificultades con estos fonemas. Sin embargo, en el análisis directo de la función de transferencia sí puede distinguirse entre resonancias y antiresonancias. La sección siguiente provee más detalles.

### 5.3. MÉTODOS PARA EL CÁLCULO DE POLOS Y CEROS

Para aproximar numéricamente los polos y ceros se recurre al método de picos [41], el cual parte de la descomposición de  $H(\omega)$  en numerador  $H_n(\omega)$  y denominador  $H_d(\omega)$ . Luego, las frecuencias

donde  $H_n(\omega) = 0$  corresponden a los ceros, y aquellas donde  $H_d(\omega) = 0$  a los polos. Así, el problema se reduce a la clásica determinación de ceros de una función mediante métodos numéricos. Como la inversión es estática, no resulta necesario computar la evolución de los polos y ceros con el tiempo.

Por supuesto, con las señales objeto no procede el cálculo de la función de transferencia. Empero, pueden aproximarse los picos espectrales aplicando Predicción Lineal (PL) de orden variable [93]. Este método aproxima el tracto vocal por medio de un modelo PL, y escruta los picos de tal modelo de acuerdo con un conjunto de reglas para confirmar su naturaleza formántica. Estos modelos PL asumen que una señal puede predecirse a partir de la combinación lineal de muestras precedentes. Aplicando el análisis PL, se computa el conjunto de coeficientes de predicción, y se usan para definir el polinomio  $P(z)$ , en la variable compleja  $z$ . Posteriormente, el filtro  $1/P(z)$ , que incluye sólo polos, funge como modelo del tracto vocal que produjo la señal analizada. Ahora, para determinar las resonancias, resulta suficiente resolver  $P(z) = 0$ . Sin embargo, debe tenerse el cuidado de rechazar los picos espurios, y de no descartar los de fina resolución. Con este fin, se aplica un refinamiento basado en reglas, donde límites en los valores de las frecuencias y anchos de banda (hasta 600 Hz), restringen los formantes admisibles, y el orden del análisis PL se incrementa buscando una mejor resolución. El filtrado de Kalman suaviza la señal [93], y una medida de continuidad espectral, basada en distancia, se utiliza para escoger el vector de formantes más representativo de la señal objeto.

## 5.4. OPTIMIZACIÓN CON ALGORITMOS GENÉTICOS CONTINUOS

El problema de optimización planteado por la Ecuación 1.1 se aborda con Algoritmos Genéticos Continuos (AGC) [24, 26, 52, 91]. En general, un algoritmo genético constituye una instancia de la familia de técnicas computacionales basadas en las nociones de la genética y la teoría de la evolución. Las soluciones potenciales de un problema específico se agrupan en *poblaciones*, y se definen operaciones análogas a los mecanismos evolutivos de *selección*, *cruce*, *mutación* y *reinserción*. La finalidad de las referidas operaciones consiste en derivar poblaciones nuevas y mejor adaptadas, desarrollando una búsqueda estocástica del espacio de soluciones. Los algoritmos genéticos resultan particularmente útiles para problemas como el de esta investigación, en los cuales el espacio de soluciones adquiere una forma irregular, con abundancia de mínimos locales: recuérdese que la inversión afronta el problema de las relaciones muchos-a-uno en la producción del habla. Otro concepto crucial es la *función objetivo*, que guía la búsqueda y cuantifica la calidad de las soluciones.

La variante continua de los algoritmos genéticos abandona la codificación binaria canónica, y la reemplaza por un gen de tipo real [68, 87, 88]. Los AGC proveen grandes beneficios para la inversión articulatoria debido a las siguientes características del problema:

1. El dominio articulatorio, y por ende el espacio de soluciones, es muy grande. En este sentido, el AGC ofrece una búsqueda amplia de la superficie de costo mediante exploraciones simultáneas.
2. No existe un método de optimización establecido para la inversión articulatoria. Los AGC optimizan variables con superficies de costo extremadamente complejas y pueden salir de mínimos locales.
3. Las variables de entrada son continuas.
4. Las interacciones entre los articuladores resultan complejas y altamente no lineales. Los mecanismos neuromusculares capaces de cambiar la forma del tracto vocal se encuentran indirectamente acoplados.

En la inversión de vocales y nasales, la distancia acústica establecida en la Ecuación 1.1 se entenderá en lo sucesivo como la distancia entre los polos y ceros de las señales objeto y de las sintéticas. Con estas últimas, en realidad, resulta innecesario desarrollar el proceso de síntesis, por cuanto los polos y ceros pueden derivarse del análisis de  $H(\omega)$ , según las funciones de área y longitud de la configuración sobre el plano medial. La síntesis se efectúa sólo al final de los experimentos de inversión, y con propósitos de validación acústica del aprendizaje.

La función objetivo formaliza la comparación de propiedades en el dominio de la frecuencia, y su forma general se deriva directamente de la Ecuación 1.1:

$$f_1(p(t), S_v) = W^T |\alpha(\phi(p(t))) - \alpha(S_v)| ./ \alpha(S_v) \quad (5.7)$$

donde ./ divide sus operandos, elemento a elemento. Naturalmente, la evaluación del vector articulatorio es relativa a la señal objeto  $S_v$ . Con el AGC la optimización procede derivando variaciones de un grupo inicial de vectores articulatorios, cuya aptitud se encuentra determinada por  $f_1$ . Cada evaluación de  $f_1$  exige en primer lugar la obtención de la configuración medial, lo que entre otras tareas requiere la invocación del sistema de inferencia difusa del Capítulo 3. Luego se calculan la función de área y la de longitud. Después de las primeras construcciones y cuantificaciones, el algoritmo ingresa a un ciclo que abandonará sólo al verificarse un criterio de convergencia preestablecido. La Figura 5.5 muestra este proceso, y las secciones siguientes proporcionan más información sobre las etapas del mismo, con las especificaciones pertinentes a los experimentos de inversión articulatoria desarrollados.

#### 5.4.1. VARIABLES Y FUNCIÓN OBJETIVO

Los niveles de actividad de los 12 músculos en  $p(t)$  representan las variables del problema. En consecuencia, cada cromosoma (potencial solución) consta de 12 genes de tipo real, cuyos valores residen en el intervalo  $\Psi$ . Esta realización concreta o codificación de un individuo se denomina *genotipo*. En un nivel superior, más vinculado a la semántica, el *fenotipo* codificado equivale a una configuración articulatoria sobre el plano medial. Por consiguiente, toda reunión de cromosomas (población) constituye un conjunto de tales configuraciones articulatorias. En resumen, la estructura del cromosoma es

$$\text{cromosoma} = p(t) = [AM_1 \ AM_2 \ \dots \ AM_{12}]$$

donde  $AM_i$  denota la actividad muscular en el  $i$ -ésimo componente del vector articulatorio. Desde otra perspectiva, la función objetivo mide el error en el que incurre un cromosoma al representar la señal objeto. De esta manera, el error asociado a un cromosoma específico, en principio, viene dado por

$$\text{error}(\text{cromosoma}, S_v) = f_1(\text{cromosoma}, S_v) = f_1([AM_1 \ AM_2 \ \dots \ AM_{12}], S_v) .$$

A esta función debe agregarse un error  $f_2$ , proveniente de la suma de todas las actividades musculares:

$$f_2(\text{cromosoma}) = \sum_{i=1}^{12} AM_i$$

La función  $f_2$  promoverá la subsistencia de aquellas configuraciones con menor gasto energético. Recuérdese que configuraciones articulatorias diferentes pueden generar señales acústicamente

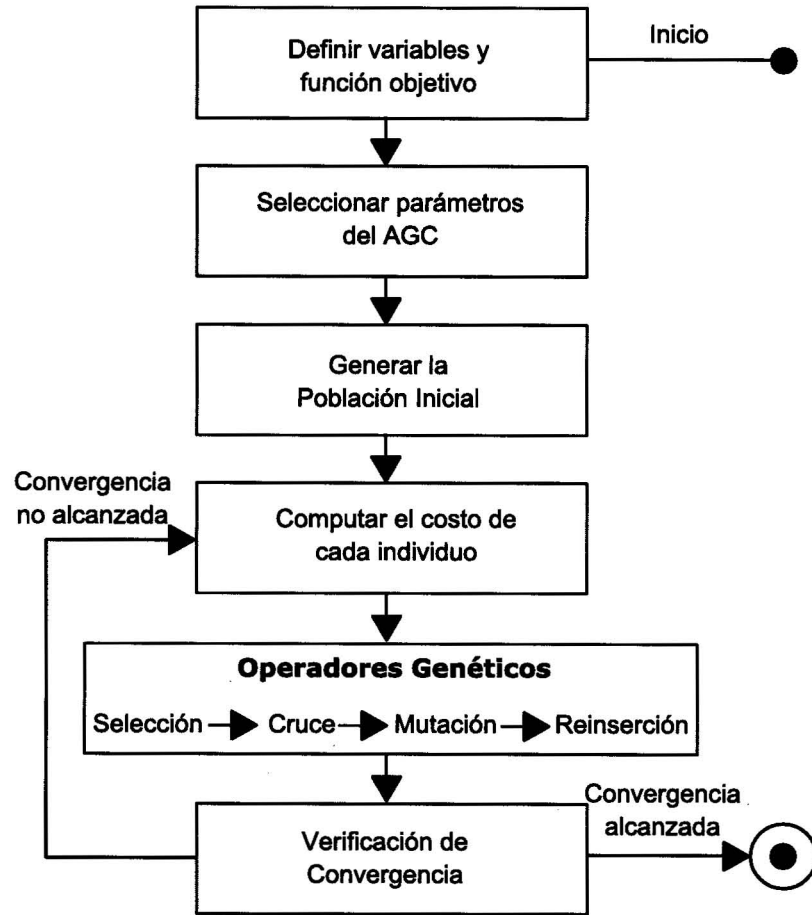


Figura 5.5: Operación general del AGC.

idénticas o muy cercanas. Por consiguiente,  $f_2$  representa el criterio para seleccionar entre las configuraciones posibles. Heurísticamente, se asumirá que la fonación prefiere las articulaciones que minimizan el esfuerzo muscular, y por consiguiente, el consumo de glucógeno. La coarticulación, por ejemplo, evidencia la tendencia del aparato fonador a producir sonidos con menor esfuerzo [66]. Al mismo tiempo, el AGC resulta consistente con la Navaja de Occam, al preferir las explicaciones más simples, que en este contexto, son las configuraciones que exigen el menor trabajo. No obstante, el mayor peso en la optimización lo tiene  $f_1$ ;  $f_2$  es sólo un auxiliar en la exploración del espacio de soluciones. En definitiva, la Ecuación 5.8 proporciona el error de un cromosoma dado.

$$\text{error}(\text{cromosoma}, S_v) = f_1(\text{cromosoma}, S_v) + f_2(\text{cromosoma}) \quad (5.8)$$

Ahora bien, la función genérica  $f_1$  de la Ecuación 5.7 debe instanciarse dependiendo del tipo de fonema a invertir. Con las vocales basta minimizar la distancia acústica de las primeras resonancias en un rango suficiente, como hasta los 5 kHz. Empero, para las consonantes nasales, la función objetivo también debe considerar la presencia de las antiresonancias. Y con las fricativas las diferencias resultan aún más drásticas. Por ende, la descripción de las instancias de  $f_1$  se posterga hasta la presentación respectiva de los experimentos de inversión efectuados. Igualmente, la determinación de los parámetros del AGC se explica en la sección 5.5.2, por cuanto depende directamente de la descripción de los experimentos.

#### 5.4.2. POBLACIÓN INICIAL

La población inicial simplemente consta de  $N_{POB}$  cromosomas cuyos genes se han obtenido aleatoriamente. La cantidad de individuos,  $N_{POB}$ , constituye uno de los parámetros determinados en la sección 5.5.2, y plantea un evidente compromiso entre la capacidad exploradora del AGC y el costo computacional de la inversión.

A cada individuo de la población se le asigna una medida de error (con la función objetivo) que será utilizada por los operadores genéticos para derivar una población nueva y probablemente mejor adaptada.

#### 5.4.3. OPERADORES GENÉTICOS

##### SELECCIÓN

La selección de individuos para el cruce sigue el esquema de Muestreo Universal Estocástico. Aquí la población se reordena al azar, y se genera un número aleatorio  $idx$  en el intervalo cerrado  $[0, \Sigma_e/N]$ , donde  $\Sigma_e$  es la suma de las evaluaciones de todos los individuos de la población, y  $N$  es la cantidad de individuos a seleccionar. A partir de  $idx$  se derivan  $N$  índices:  $idx, idx + (\Sigma_e/N), \dots, idx + (\Sigma_e/N)(N - 1)$ . Estos índices apuntan a  $N$  individuos en el intervalo de evaluación acumulada de toda la población, y así los individuos más aptos tienen mayor probabilidad de resultar escogidos para el cruce. Esta clase de muestreo posee una complejidad algorítmica lineal, con una implementación más simple que la selección con ruleta. En los experimentos subsiguientes se ha utilizado  $N = N_{POB}$ .

##### CRUCE

Los genes  $O_i$  del nuevo individuo obtenido del cruce de dos progenitores ( $P_1$  y  $P_2$ ) vienen dados por la Ecuación 5.9 [47, 49], donde  $s_c$  es una constante de escalado en el intervalo cerrado  $[-0.25, 1.25]$ .



$$O_i = P_{1,i} + s_c(P_{2,i} - P_{1,i}) \quad (5.9)$$

$s_c$  se escoge aleatoriamente para cada gen de los progenitores. Geométricamente, este tipo de cruce genera nuevas variables en un hipercubo de dimensiones ligeramente superiores a las de aquellos definidos por los padres, pero restringido por el valor de  $s_c$  [49]. Si el nuevo gen reside fuera del intervalo  $\Psi$  se repite el cálculo luego de seleccionar otro  $s_c$ .

### MUTACIÓN

La mutación constituye el operador más simple de este AGC. Por cada gen se genera un número aleatorio entre 0.0 y 1.0, y si supera la probabilidad global de mutación  $M_r$ , el gen se reemplaza por un número aleatorio en  $\Psi$  [24]. Teóricamente, la mutación garantiza que toda región del espacio de soluciones tiene alguna probabilidad de ser explorada. El valor de  $M_r$  en ocasiones resulta crítico para la velocidad de convergencia de los algoritmos genéticos continuos [24], por lo que se ha preferido establecerlo a partir de una inversión preliminar en la sección 5.5.2.

### REINSERCIÓN

Los nuevos individuos deben introducirse en la población reemplazando a los menos aptos. Aquí usaremos una inserción *basada en ajuste*, donde la nueva población se conforma seleccionando los  $N_{POB}$  individuos más aptos del conjunto integrado por la población actual y el 90% de los descendientes (los que exhiban el menor error). Esa razón de 0.90 se denomina *brecha generacional*, y garantiza que el individuo más apto de la generación previa pase a la siguiente, lo cual en el fondo implementa una estrategia elitista. El elitismo, nuevamente, persigue una mayor velocidad de convergencia [24].

#### 5.4.4. VERIFICACIÓN DE CONVERGENCIA

En la Figura 5.5, la salida del ciclo ocurre al lograrse la convergencia. Concretamente, el AGC se detendrá si alguna de las siguientes dos condiciones se verifica:

1. Umbral de error: El mínimo valor de  $f_1$  es cero, pero obviamente el de  $f_2$  depende de  $S_v$ , y con seguridad superará a cero. Como ignoramos a priori la actividad muscular involucrada en las pronunciaciones, el umbral del máximo error tolerable se igualará a 12.0, máximo valor de  $f_2$ . Una prueba preliminar en el modelo articulatorio revela que la configuración con  $f_2 = 12.0$  es anatómicamente inviable, y por ende el AGC no logrará la convergencia con  $f_1 = 0$  y  $f_2 = 12.0$ . Se espera que si el error desciende a 12.0, la configuración recuperada resulte próxima, mas no idéntica, a la señal objeto (criterio de  $f_1$ ) y con relativamente poco gasto energético (criterio de  $f_2$ ).
2. Cota de generaciones: Se establece en 80 como máximo. Aunque este valor es arbitrario, los experimentos de inversión, como se verá, suelen converger en un número mucho menor de generaciones.

En caso de no satisfacerse los criterios, se evalúa la aptitud de la nueva población y se reitera el proceso, hacia la próxima generación.

Una vez definidos los operadores genéticos y el criterio de convergencia del AGC, se procede al desarrollo de los experimentos de inversión. Específicamente, el AGC invierte un grupo de señales

objeto, grabadas a locutores adultos. Cada inversión debe proporcionar un vector articulatorio que permita generar una señal sintética acústicamente similar a la señal invertida, utilizando los modelos articulatorio y acústico.

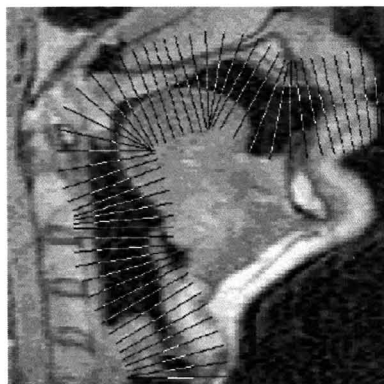
## 5.5. EXPERIMENTOS DE INVERSIÓN ARTICULATORIA

### 5.5.1. CORPUS DE SEÑALES OBJETO

El corpus agrupa las señales objeto a invertir, muestreadas a 16 kHz, con factor de cuantización de 16 bits. Esta frecuencia de muestreo asegura una banda de 8 kHz, suficiente para la caracterización de las cinco vocales españolas, las consonantes nasales /m/ y /n/, y las consonantes fricativas /f/ y /s/. Las señales objeto se extrajeron de grabaciones de 4 locutores adultos (2 masculinos y 2 femeninos), con tonos diferentes. Cada locutor produjo una grabación por cada fonema. Sin más identificación, los locutores masculinos se etiquetaron como  $M_1$  y  $M_2$ , y los femeninos,  $F_1$  y  $F_2$ . En otras investigaciones, la inversión ha utilizado una sola señal objeto [16], y en otras hasta 6 locutores [72–74,76]. Como se invertirán 9 fonemas, el total de grabaciones asciende a 36, 18 de voces masculinas, y 18 de voces femeninas. Algún preprocesamiento sobre estas señales resultó necesario para obtener las señales objeto definitivas. En particular, se suprimieron los silencios al final y comienzo de las grabaciones de vocales. Para las consonantes se grabaron las secuencias *ma*, *na*, *fa* y *sa*. La terminación de las sílabas en /a/ minimiza los efectos perturbadores de la coarticulación<sup>2</sup>, por cuanto /a/ es la vocal más central. En este caso, además de la supresión de silencios, se retiró la porción de señal correspondiente a la vocal /a/ y la transición.

### 5.5.2. INVERSIÓN DE LA VOCAL /u/

El primer experimento invierte la vocal /u/. Esta vocal requiere un desplazamiento importante de la masa lingual en dirección del proceso estiloide, aunado al abocinamiento de los labios formando un tubo acústico anterior (Figura 5.6). De este modo, la “distancia” entre la posición de equilibrio y la configuración articulatoria correcta resulta significativa, lo cual convierte a esta vocal en un objetivo idóneo para verificar, en una primera aproximación, la efectividad del AGC. Por tal razón, los parámetros empleados en el resto de experimentos se derivan de esta inversión.

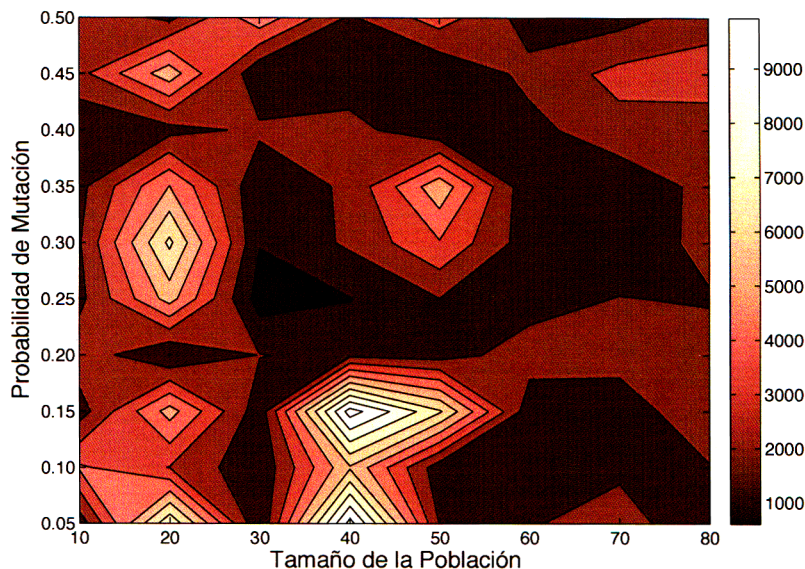


**Figura 5.6:** Resonancia magnética de la vocal /u/ [35].

<sup>2</sup>La coarticulación complica la *inversión* articulatoria. Ahora, los *synetizadores* articulatorios deberían reproducir sin problemas los efectos de la coarticulación.

Con cualquier vocal la función objetivo promoverá la subsistencia de aquellas configuraciones cuyas primeras cuatro resonancias aproximen los picos espectrales respectivos de la señal objeto, lo que equivale a favorecer la predominancia de los individuos con estructura formántica similar a la de  $S_v$ . Luego,  $p(t) = cromosoma$ , y  $\alpha(\phi(p(t)))$  devuelve las 4 primeras resonancias vinculadas a la configuración articulatoria  $cromosoma$ . En el cálculo de  $\alpha(\phi(p(t)))$  en realidad no se desarrolla la síntesis denotada por  $\phi$ . Mucho más eficiente resulta derivar la configuración medial codificada en el cromosoma, a partir de la cual se obtienen las funciones de área y longitud requeridas para el cómputo de la función de transferencia. La sección 5.2 explicó la manera de calcular las resonancias a partir de la función de transferencia. A su vez,  $\alpha(S_v)$  retorna los formantes F1-F4 de la señal objeto, a través del análisis PL. Luego, la evaluación que  $f_1$  hace del  $cromosoma$ , viene dada por la Ecuación 5.7, con  $W = [100\ 60\ 30\ 15]^T$ . Los coeficientes de cada término del error reflejan la importancia acústica de cada formante. Finalmente,  $error(cromosoma, S_v)$  tiene la forma de la Ecuación 5.8, anexando la métrica de actividad muscular. En la práctica, no todo cromosoma implica una configuración medial anatómicamente viable. Dependiendo de los valores de las variables, pueden producirse configuraciones con fronteras superpuestas, o con área igual a cero en alguna sección. Cuando se detecte uno de tales cromosomas la evaluación se apartará de la Ecuación 5.8, y se establece directamente  $error(cromosoma, S_v) = 1500$ , para forzar la remoción de dicho individuo en las próximas generaciones.

Ahora se obtiene una gráfica con el promedio de evaluaciones de la función objetivo, en 40 ejecuciones del AGC (10 por cada señal /u/ del corpus), para diversos tamaños de la población y probabilidades de mutación. La zona más oscura de la Figura 5.7 denota la menor cantidad de evaluaciones, y en consecuencia, se fijan  $N_{POB} = 35$  y  $M_r = 0.26$  para todos los experimentos subsiguientes.



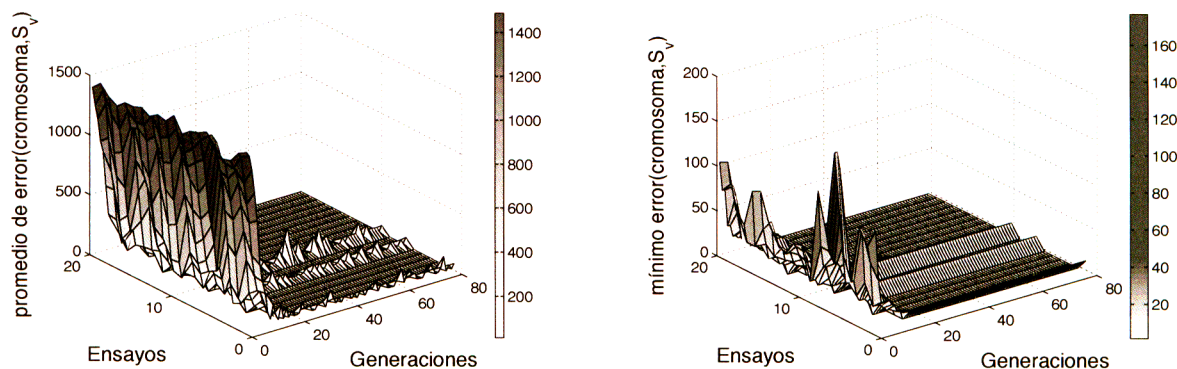
**Figura 5.7:** Gráfica de contorno para la obtención de parámetros del AGC.

Una vez completados los parámetros del AGC, se procedió a la inversión formal de la /u/, aplicando 20 ensayos a cada instancia de esta vocal en el corpus. Más adelante, el Cuadro 5.1 recopila los resultados del mejor ensayo, por locutor. Sin embargo, antes se muestra la evolución de la función objetivo y las mejores configuraciones articulatorias en las Figuras 5.8 y 5.9, respectivamente, provenientes de los ensayos donde se obtuvo el menor error, distinguiendo entre locutores masculinos

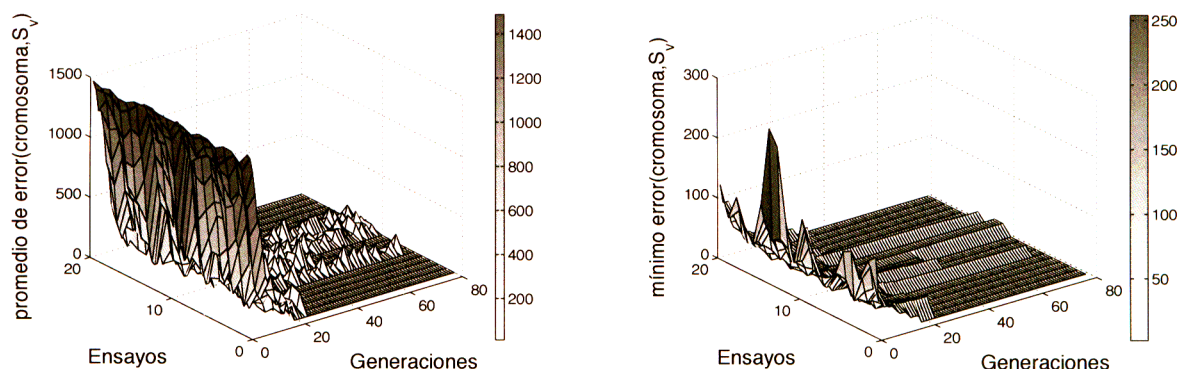


y femeninos. En este sentido, los locutores con el menor error en la inversión fueron  $M_2$  y  $F_1$ . La Figura 5.8 muestra el curso del promedio y el mejor valor de la función objetivo durante los ensayos de inversión de estos locutores, y como se aprecia, la convergencia ocurre rápidamente.

(Locutor  $M_2$ )



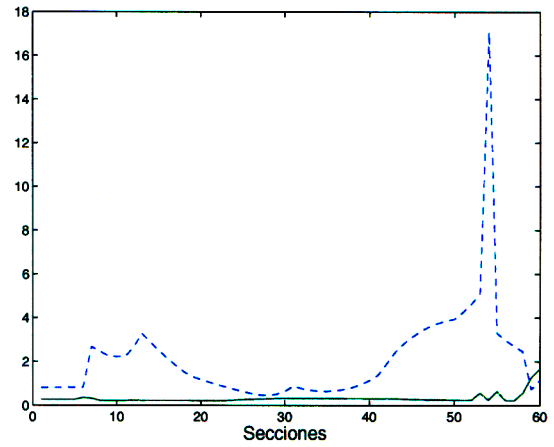
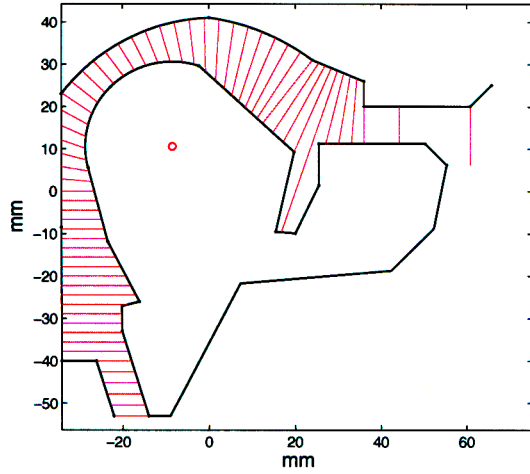
(Locutor  $F_1$ )



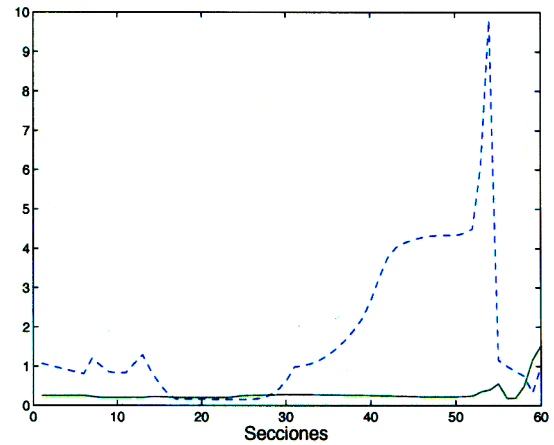
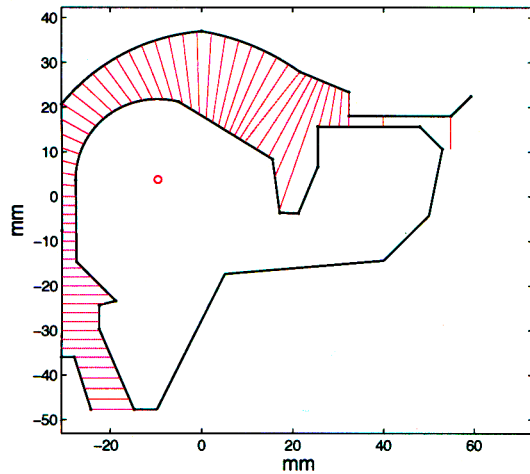
**Figura 5.8:** Promedio y mejor valor de la función objetivo por generación (vocal /u/).

La Figura 5.9 exhibe las mejores configuraciones articulares mediales correspondientes a los experimentos en la Figura 5.8. Soslayando las inevitables limitaciones del modelo medial, las configuraciones revelan proximidad a la resonancia de la Figura 5.6, manifestándose la retracción de la masa lingual y la conformación en los labios de un puerto de radiación más largo. Estos resultados demuestran que, en general, la inversión resultó exitosa. Sin embargo, un examen más cuidadoso encuentra algunos detalles interesantes, específicamente al analizar los vectores articulatorios de la Figura 5.10. Como se esperaba, el OO interviene en gran medida. Además, como las vocales constituyen eventos acústicos con un tracto vocal relativamente abierto, el MH debe actuar, como correctamente demuestran los vectores. Empero, la retracción de la lengua en ambas configuraciones apela a distintos mecanismos. En el caso masculino, el GGp eleva la masa lingual, el GGa compensa la elevación en el ápice, y el SG completa la retracción con una leve participación. Resulta difícil establecer la exactitud fisiológica de este vector articulatorio, ante las limitaciones del modelo articulatorio, y por la inexistencia de la base de datos pertinente a la actividad muscular real. No obstante, el AGC ha logrado recuperar una espléndida configuración articular utilizando la información disponible. Ahora bien, la configuración femenina evidencia el problema intrínseco al

(Locutor  $M_2$ )



(Locutor  $F_1$ )



**Figura 5.9:** Mejores configuraciones articulatorias recuperadas (vocal /u/). A la derecha de cada configuración medial se muestran las funciones respectivas de área (punteada, en  $\text{cm}^2$ ) y longitud (continua, en  $\text{cm}$ ).