

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

**Producción de la voz, control
neuronal y biometría**

por Marcos Alberto Trevisan

Director de Tesis: Dr. Gabriel Mindlin.

Lugar de Trabajo: Departamento de Física, FCEyN, UBA.

Trabajo de Tesis para optar por el título de
Doctor de la Universidad de Buenos Aires en el área de Física

22 de mayo de 2006

A mi padre.

Índice general

Resumen	7
Abstract	9
delineamientos	13
1. Introducción	15
1.1. La producción de la voz	15
1.1.1. Anatomía y biomecánica	15
1.1.2. Representación del habla en el tiempo y en la frecuencia	19
1.2. Técnicas de reconstrucción espectral	22
1.3. Biometría	25
1.3.1. Principios de reconocimiento de identidad por la voz	26
2. Física de la producción de la voz	29
2.1. Modelo de Fuente-Filtro	30
2.1.1. Fuente: modelos de cuerdas vocales	30
2.1.2. Filtro: el tracto vocal	35
2.2. Ajuste de parámetros	37
2.2.1. Algoritmos genéticos	37
2.2.2. Adaptación al problema de la producción de voz	38
2.3. Resultados	39
3. Sistemas dinámicos y topología	43
3.1. Sistemas dinámicos forzados	43
3.2. Invariantes topológicos	45
3.2.1. Tasas de rotación relativa	46
3.2.2. Números de conexión	47
3.2.3. Series embebidas	48
4. Huellas vocales y biometría	51
4.1. Topología y voz	52
4.1.1. Reconstrucción espectral de sonidos voceados	52
4.1.2. Construcción de matrices topológicas	54
4.1.3. Índices y órbitas	55

4.1.4. Implementación del método topológico	57
4.2. Resultados y discusión	57
5. Control neuronal vocal	61
5.1. Actividad neuronal	62
5.2. Neuromecánica vocal	65
5.2.1. Siringe y glotis	65
5.2.2. Pulmones y sacos aéreos	67
5.3. Subarmónicos y complejidad vocal	68
5.3.1. Conclusiones	74
5.4. Control neuronal y aprendizaje vocal	75
5.4.1. Soluciones e interpretación	76
5.4.2. Aplicación a modelos de actividad promedio	80
5.5. Discusión	82
A. Algoritmos genéticos	85
A.0.1. Adaptación del algoritmo a un problema matemático simple	85
A.0.2. Bases teóricas de los algoritmos genéticos	86
A.1. Paralelización del algoritmo	89
B. Métodos	91
B.1. Medición de la presión de sacos aéreos	91
Bibliografía	93
Agradecimientos	97

Resumen

Este trabajo puede separarse en dos partes. La primera está dedicada al problema biométrico de la voz, es decir, al uso de la voz como caracterización unívoca del hablante. La segunda está dedicada al control neuronal vocal.

En la historia de la biométrica, los intentos de descifrar la forma en que está codificada la identidad de una persona en su voz están basados en tratamientos estadísticos de sus propiedades espectrales. El éxito relativo de estas técnicas se debe a la dificultad de establecer distancias entre estas propiedades. Todos los sistemas de identificación necesitan umbrales de decisión para aceptar o rechazar usuarios.

En la primera parte exploramos dos alternativas a este punto de vista. En primer lugar, se contruyó un modelo de producción de voz de baja dimensión con el objeto de reproducir la diversidad de los procesos vocales y poder describirlos en términos de un mínimo conjunto de parámetros anatómicos relevantes. Una vez construido el modelo, una pregunta surge inmediatamente: ¿Podrán ajustarse sus parámetros para identificar la voz de cada persona? Los intentos por responder esta pregunta inauguran la segunda alternativa, que consiste en la definición de *huellas vocales topológicas* basadas en herramientas creadas para caracterizar la estructura de las órbitas de los sistemas dinámicos, y que nos permitieron traducir cualidades espectrales de la voz en índices topológicos robustos. Los resultados sugieren una nueva dirección en la identificación de las personas por la voz a través de arreglos de números enteros, independientes de umbrales de decisión.

La segunda parte de este trabajo está dedicada a la interacción del sistema vocal de un modelo animal sencillo (aves) con el sustrato neuronal que lo controla.

En las últimas décadas, el estudio del canto de las aves se desarrolló hasta convertirse en un campo de investigación privilegiado, permitiendo la experimentación con organismos que desarrollaron habilidades de aprendizaje vocal con fuertes paralelos con la adquisición del habla.

Este campo se está transformando también en un área de modelado neuronal. La investigación en este campo busca entender la emergencia de patrones acústicos complejos y la estructura del aprendizaje vocal, independiente de factores culturales y sociales. Basándonos en el modelo vocal desarrollado para el habla, la segunda parte del trabajo está dedicada a construir modelos neuronales para aproximarse a estos dos problemas.

Palabras clave: física de la voz, biometría, topología, control neuronal vocal.

Abstract

This work is twofolded. The first part is devoted to voice biometrics, i.e. the problem of a one to one characterization between voice and speaker. The second part is dedicated to the neural vocal control.

In the history of voice biometrics, efforts to unveil the speaker's identity were based on spectral descriptions of voiced sounds. The relative success of these techniques is due to the difficulty of establishing how close is close enough when comparing spectral features. Every system needs a threshold to be defined in order to accept or reject users.

In the first part of this work we explore alternatives to this point of view. In the first place, we built a low dimensional dynamical system which allowed us reproducing the diversity of the vocal processes in terms of a minimal set of anatomically relevant parameters. Once the model was constructed, a question naturally arised: Could its parameters reproduce a particular speaker's voice? Our results leded to a second perspective which introduces the concept *topological voiceprints* based on a set of tools created in order to characterize the orbits structure of dynamical systems. That tools allowed us to translate spectral features into sets of robust integers. Our results suggest a new direction in the identification of subjects by voice: one in which arrangements of integers define voiceprints that stand on their own, despite any acceptance/rejection thresholds.

The second part of this work is devoted to the interaction of the vocal system in a simple animal model (birds) and the neural substrate controlling its dynamics.

During the last few decades, birdsong developed into a privileged research field that allowed neuroscientists to experiment with biological organisms that developed vocal learning skills presenting strong analogies with language acquisition.

In the present, birdsong is becoming a definite field of neural modeling. Research in this field is devoted to understand the emergency of complex acoustic patterns and to study the neural architecture of vocal learning, independently from social and cultural factors involved in language acquisition. Based on the human vocal model, we developed neural models to address these two specific problems.

Keywords: voice physics, biometrics, topology, neural motor control.

Motivación y estructura

Motivación y objetivos

El estudio de la voz humana presenta un doble desafío. Por un lado, el problema teórico en el que convergen distintas áreas de la matemática y de la física. La mecánica de los fluidos, la teoría de elasticidad, la acústica y la dinámica no lineal buscan enmarcar la enorme variabilidad acústica en términos de flujos de aire laminares o turbulentos, acoplamientos entre la fuente de sonido y su ‘caja resonante’, distintos modos de oscilación de las cuerdas vocales. En este contexto, el objeto más importante del estudio de la voz es la captura de los principios que regulan la diversidad de sonidos del instrumento vocal, y el estudio de sus ingredientes anatómicos y fisiológicos esenciales.

Por otra parte, la naturalidad en la percepción del habla resulta muy persuasiva para el estudio del reconocimiento de la voz. La identificación de los sonidos del habla y de las personas por su voz nos resulta tan natural y evidente que tendemos a desvincularlos del resto de los sonidos. Seducida por esa naturalidad, la curiosidad propone una serie de preguntas acerca de la forma en que la voz codifica la información del aparato vocal que la produce. Esas preguntas se traducen sometiendo los modelos teóricos de producción de la voz a tareas de reconocimiento. Las aplicaciones tecnológicas actuales no están basadas en la física de la fonación, sino vagamente inspiradas en una de las funciones del oído interno, que consiste en caracterizar los sonidos espectralmente. El vacío que queda entre los modelos basados en la física de la voz y la descripción meramente estadística de los sistemas de reconocimiento abre una gran brecha para el estudio del problema.

Más allá de la mecánica de la audición, las operaciones de procesamiento neuronal exigen un campo específico de estudio, que en los últimos años cobra cada vez una identidad más definida. El sistema neuronal auditivo alimenta al sistema motor, corrige errores, y activamente controla la producción de la voz. En particular, el modo en que el sustrato neuronal gobierna y controla al aparato vocal es un problema central de la neurociencia aplicada al campo de la voz. El resultado de la ejecución de comandos neuronales es una inmensa variedad de respuestas motoras que permiten a los individuos interactuar entre sí y con un ambiente en permanente cambio.

El sistema nervioso está formado por enormes redes neuronales condensadas en estructuras de neuronas densamente interconectadas (núcleos), que se proyectan a su vez a otros núcleos a través de interneuronas. El primer problema que enfrenta entonces el modelado de los sistemas neuronales es el de la escala de descripción del problema. Por un lado, experimentos con neuronas individuales muestran patrones complejos de actividad. Por otra parte, la organización jerárquica de los núcleos neuronales sugiere una visión más global del problema, en el que la actividad conjunta de miles de neuronas condensadas en cada núcleo impone su actividad.

¿Es la diversidad de respuestas el resultado de la interacción de un sistema neuronal intrínsecamente extenso? ¿Existen propiedades globales que puedan explicarla? Estas son preguntas abiertas de la neurociencia en general, y en este trabajo estudiamos los alcances de la segunda perspectiva aplicados a la producción vocal.

Casi todas las aves comparten ciertos aspectos básicos de la fonación con los humanos. Por ejemplo, el aparato fonador está controlado por los mismos mecanismos físicos. Otros grupos específicos de aves, los *oscinos*, presentan además fuertes paralelos con ciertos aspectos de la adquisición del habla. Todas estas características hacen de este campo uno ideal para la construcción de modelos que pueden ser contrastados anatómicamente y fisiológicamente, en un sistema biológico que puede ser estudiado *in vivo*, interactuando con su ambiente.

La interacción de los órganos periféricos con el cerebro y la generación de comandos motores que dan origen a patrones acústicos complejos cierra una variedad de preguntas que animaron el trabajo compilado en esta tesis.

Organización del trabajo

Este trabajo está organizado en cinco capítulos.

En el primero presentamos los elementos anatómicos esenciales de la producción de la voz y las herramientas tecnológicas asociadas al reconocimiento de la identidad, tal como se las concibe actualmente.

El segundo capítulo está dedicado a la fisiología vocal y a los modelos de la física de producción de la voz. Allí estudiamos los procesos elementales que dan cuenta de la dinámica de las cuerdas vocales y la generación de los sonidos del habla. Con el espíritu de vincular la identificación de la identidad con el aparato vocal, sometimos el modelo construido a un método de identificación por la voz.

Los resultados de esta sección sugieren una nueva aproximación al problema, basado en un programa diseñado para clasificar y sistematizar topológicamente la estructura de órbitas periódicas de sistemas dinámicos que presentamos en el capítulo tercero. Además, se presentan algunas propiedades de los sistemas dinámicos forzados, que serán usadas para modelar el control neuronal vocal al final de este trabajo.

En el cuarto capítulo se define el concepto de *huella vocal topológica*, que

permite una aproximación al problema del reconocimiento de la identidad por la voz que es cualitativamente distinto a la tecnología existente, independiente de umbrales de decisión. En este capítulo se exportan las herramientas presentadas en el capítulo anterior, y se analiza una implementación topológica para la identificación de personas por su voz. Los resultados y las direcciones futuras de investigación se discuten al final del capítulo.

El quinto capítulo se dedica al modelado del control vocal neuronal bajo la perspectiva de la actividad media de los núcleos motores que controlan los gestos fundamentales de la fonación: la presión pulmonar y la configuración de las cuerdas vocales para la fonación.

Proponemos aquí una descripción de actividad neuronal media para estudiar el nivel de emergencia de la complejidad acústica en distintos grupos de aves.

Capítulo 1

Introducción

‘Tal vez será su voz, tal vez...’
Manzi, Demare.

En este capítulo se discuten principios de la producción y percepción de la voz y se revisan las técnicas espectrales más típicas en las áreas del reconocimiento de la voz y de la identidad a través de la voz.

1.1. La producción de la voz

1.1.1. Anatomía y biomecánica

La anatomía del sistema vocal completo es compleja. Funcionalmente, además de las cuerdas vocales contenidas en la *laringe*, el sistema está formado por las vías respiratorias (los *pulmones*, *bronquios*, *tráquea*, *faringe*, *cavidad oral* y *nasal*), la cabeza y el cuello. En esta sección se propone una breve introducción a los elementos vitales de la anatomía que resultarán caricaturizados en los modelos matemáticos de producción de la voz.

En el sentido del flujo de aire, estos elementos son: el *sistema pulmonar*, que es la fuente del flujo de aire, la *laringe*, donde se alojan las cuerdas vocales que oscilan y el *tracto vocal/nasal*.

Sistema Pulmonar

Por debajo de la laringe, el sistema respiratorio se divide en *alvéolos* y *vías respiratorias*. Para la actividad vocal, la función principal de este sistema es la de proveer flujo de aire a través de la laringe, de modo que las cuerdas vocales alojadas allí puedan oscilar.

Las vías respiratorias se internan en los *pulmones*, que definen regiones confinadas que contienen aire presurizado. Cada uno de los pulmones está conformado por millones de alvéolos interconectados que se acortan y estrechan a medida que se extienden más profundamente en el tejido pulmonar, para terminar en

los alvéolos. Los *bronquiolos terminales* representan el punto de mayor profundidad, a través donde el intercambio de gases se produce por difusión simple a favor del gradiente de concentración. El propósito de esta estructura ramificada es el de aumentar el área de contacto entre el tejido y el aire.

El *tórax* consiste en una serie músculos, tejido conectivo y huesos, entre ellos las costillas, cerradas por detrás por la columna vertebral y por delante por el esternón, creando así una caja de protección que envuelve a los pulmones y al corazón. Cerrando esta caja por debajo, se encuentra una estructura de músculos y tendones llamada *diafragma*, que separa el tórax de las vísceras. El diafragma, según se verá a continuación, juega un rol fundamental en el mecanismo respiratorio.

Los pulmones están sellados doblemente dentro de la caja torácica. La membrana interna se llama *pleura visceral*, y a la externa, *pleura parietal*. El espacio interpleural está relleno de líquido, permitiendo el deslizamiento de las membranas y facilitando el movimiento de expansión y compresión pulmonar.

Durante una conversación normal, la presión pulmonar p_l varía típicamente entre 0.3 y 1.2kPa, resultando una pequeña fracción de la presión atmosférica. Este es un hecho afortunado desde el punto de vista del modelado porque permite simplificaciones respecto de la presión acústica en el tracto vocal (capítulo 2).

Para relacionar volumen y presión pulmonar, se observa la validez de la ecuación de los gases ideales $p_l V = nRT$ a temperatura constante. Cuando el diafragma se deforma y baja, la caja torácica se expande y el volumen pulmonar aumenta. La presión pulmonar p_l disminuye y, una vez abierta la glotis, el aire fluye hacia el interior de los pulmones. Inversamente, la restitución elástica de los pulmones y la contracción de los músculos intercostales internos provoca una disminución del volumen pulmonar, creando un gradiente de presión que expelle el aire hacia el exterior.

El tórax y abdomen, junto con los pulmones, actúan como fuelles que producen un *flujo de aire* cuya intensidad es regulada por las constricciones de las vías respiratorias superiores; durante el habla, el control de estas constricciones en la laringe o en el tracto vocal permite generar una amplia variedad de sonidos distintos.

Además de la revisión de la anatomía relacionada con la actividad vocal, necesaria para el desarrollo de los modelos de fonación, en el capítulo final revisaremos el control vocal neuronal. Para entonces interesará también un breve compendio del funcionamiento de los centros neuronales respiratorios.

La respiración es controlada por un generador de actividad neuronal llamado el *centro de respiración*, constituido por tres grandes grupos de neuronas en la parte baja del cerebro. El *grupo respiratorio dorsal* es predominantemente responsable de la inspiración. El *grupo respiratorio ventral* puede causar inspiración y expiración, y el *centro neumotorácico* ayuda a controlar ritmos y patrones respiratorios.

Mientras que la inspiración es activamente controlada por centros neuronales que gobiernan los músculos intercostales externos, la expiración resulta princi-

palmente del efecto mecánico de la restitución elástica de los pulmones y la caja torácica.

Además, localizados en las paredes de los bronquios y bronquiolos existen receptores de estiramiento ligados a los grupos neuronales de la respiración que inhiben la actividad de inspiración cuando la expansión pulmones supera ciertos límites.

El movimiento de los pulmones es controlado por el *diafragma*. Durante la expiración, la dinámica del diafragma puede ser modelada a través de la ecuación [Kee98]

$$m\ddot{x} + \kappa x + \mu\dot{x} = f(\alpha_1, \alpha_2, \dots, \alpha_n) \quad (1.1)$$

donde f es una función de las *actividades neuronales* α_i de los n grupos neuronales distintos que enervan los músculos respiratorios.

El concepto de actividad neuronal se desarrollará con mayor profundidad en el capítulo 5. Aquí bastará aclarar que cada neurona puede caracterizarse por una traza de pulsos eléctricos (potenciales de acción) estereotipados que corresponden a estímulos por encima de un umbral. La actividad α_i resulta de la integración de las actividades individuales que forman los grupos neuronales densamente interconectados (núcleos) que enervan los músculos.

El diafragma queda así modelado por una masa m amortiguada cuyo desplazamiento medio x es gobernado por núcleos neuronales específicos y está sujeto a una fuerza restitutiva de coeficiente κ y disipación caracterizada por el coeficiente μ .

Laringe

La laringe es un sistema de cartílagos y músculos que contiene el tejido blando que compone las cuerdas vocales. En la figura 1.1 se ve un esquema de sus principales componentes: de abajo hacia arriba, los cartílagos *cricoide* y *tiroide*, y el hueso *hioide*, unidos por ligamentos. Un sofisticado sistema muscular, que no está esquematizado, interconecta los cartílagos y huesos, permitiendo una gran flexibilidad a la laringe en sus funciones vocales, de deglución y ventilación. En la sección más angosta de la laringe (*glotis*) se localizan las cuerdas vocales. Hacia arriba, el ventrículo laríngeo y luego las cuerdas ventriculares (o falsas cuerdas). Finalmente, los pliegues ariepiglóticos. En combinación, el sistema de cuerdas y pliegues pueden ocluir completamente la vía laríngea.

Las cuerdas vocales están constituidas esencialmente por una estructura conocida como *lámina propia*, que es un sistema de capas de tejido no muscular. Hacia las capas internas del tejido, se encuentra la *mucosa*, que incluye el epitelio y la capa más superficial de la lámina propia, el *ligamento*, formado por las capas intermedia y profunda de la lámina propia, y el músculo *tiroartenoideo*. Esta microestructura de capas consiste esencialmente en una fuerte concentración de fibras, que no están presentes solamente en la porción muscular sino en las varias capas de la lámina propia.

Durante el habla, con la laringe en posición de fonación, las cuerdas vocales no varían demasiado su configuración, manteniendo una frecuencia fundamental

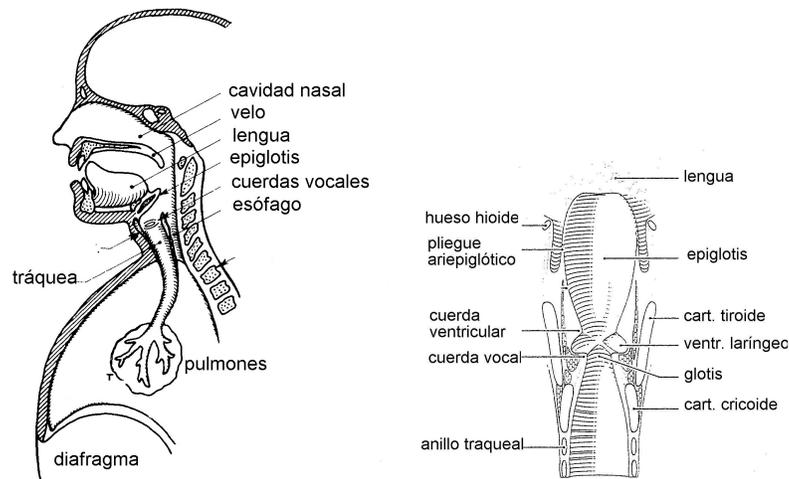


Figura 1.1: Vista esquemática de los elementos comprometidos en el sistema vocal/respiratorio (izquierda) y los de la laringe, donde se alojan las cuerdas vocales.

de oscilación más o menos constante. Aun en estas condiciones de frecuencia fundamental aproximadamente constante, su control es importante para definir el contorno melódico del discurso [Tit93], en el que interviene una sofisticada red de músculos que cambian la geometría de la laringe, modificando la tensión de las cuerdas vocales y con ella la frecuencia fundamental de las oscilaciones. La gran variedad de configuraciones musculares puede generar redundancias [Tit93]. Por ejemplo, cuando el músculo *cricotiroideo* (CT) está contraído y el *tiroaritenoides* (TA) está relajado (inactivo), las cuerdas vocales se estiran, aumentando su tensión, e incrementando la frecuencia fundamental. Por otra parte, si CT y TA se contraen simultáneamente, el resultado es también un incremento de la frecuencia fundamental.

Tracto vocal y nasal

Por arriba de la laringe, el sistema de respiración/fonación está constituido por dos series de cavidades: el *tracto vocal* y el *tracto nasal*.

El tracto vocal está formado por la cavidad que se extiende desde la laringe, a la salida de las cuerdas vocales, hasta los labios. Está formado por la *faringe*, que conecta el esófago con la boca y la *cavidad oral*, con una longitud promedio de 17,5 cm en los varones adultos y de 15,5 cm en mujeres adultas. La sección transversal a lo largo del tracto vocal queda determinada por las posiciones de la lengua, labios, mandíbula y el velo, y puede variar entre 20 cm² y cero, correspondiente al cierre completo.

El tracto nasal comienza en el velo y termina en las fosas nasales. Cuando el velo baja, actúa como una válvula que acopla acústicamente los dos tractos

para producir los sonidos nasales del habla.

Con los elementos anatómicos descritos, un esquema del mecanismo vocal puede trazarse. El aire entra a los pulmones mediante el mecanismo normal de respiración. Conforme el aire es expelido a través de la tráquea, las cuerdas vocales tensas en la laringe pierden su estabilidad y entran en un régimen de oscilaciones sostenidas usando parte de la energía del flujo de aire para compensar las pérdidas por disipación. A partir de entonces, las fluctuaciones de presión viajan por el tracto vocal, que presenta una sección variable, a la salida del cual un micrófono registra serie temporales como las que se muestran en las figuras 1.2 y 1.3(a).

Antes de modelar los procesos descritos en este capítulo, nos detendremos en las distintas representaciones de la señal acústica que serán útiles para las próximas discusiones.

1.1.2. Representación del habla en el tiempo y en la frecuencia

La serie temporal del habla registrada por un micrófono muestra una señal que varía lentamente en el tiempo cuando es examinada en períodos de tiempo $t \in (5, 100)ms$. En ese rango temporal la señal se mantiene predominantemente estacionaria, coincidiendo con la duración de los diferentes fonemas durante el habla normal. En los paneles de la figura 1.2 se ven los segmentos de una serie temporal correspondiente a los fonemas /o/ y /e/, y a la consonante /s/ extraídas de la palabra *voces*, en ventanas de alrededor de 100 ms. Para tiempos mayores que 1/5 s, sin embargo, la serie temporal cambia, dando cuenta de los distintos sonidos pronunciados (figura 1.3).

Sonograma

Además de la serie temporal, una alternativa muy útil para caracterizar y extraer información de la señal del habla es el *sonograma* o *espectrograma*, que representa su contenido espectral a lo largo del tiempo. Para eso, se hace corresponder una ventana de análisis a cada tiempo de muestreo sobre la que se calcula el espectro, que se representa por una línea vertical que recorre las frecuencias analizadas, representando mayor carga espectral con un color más oscuro. El ejemplo de la figura 1.3 para la palabra *voces* aclarará el procedimiento. Naturalmente, el rango de frecuencias analizado se corresponderá con la resolución temporal Δ de la serie original, con frecuencia máxima $f_{max} = 1/2\Delta$. La resolución temporal del espectrograma quedará definida por la ventana temporal de análisis: a mayor tamaño de ventana de análisis, mayor resolución espectral al costo de una resolución temporal más pobre.

Clasificación de los sonidos del habla

Algunos sonidos del habla admiten una primera distinción, que ya se presentó, entre aquellos originados mediante la excitación de un tracto vocal esen-

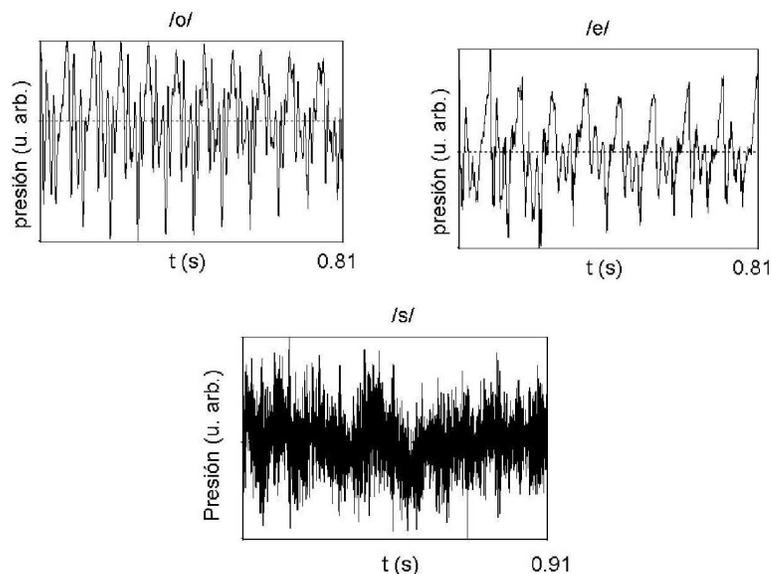


Figura 1.2: Segmentos extraídos de la enunciación de la palabra ‘voces’ (figura 1.3). Cada segmento se mantiene aproximadamente estacionario durante $\Delta t \sim 100\text{ms}$.

cialmente fijo por las cuerdas vocales oscilantes (*voceados*), y los originados en la turbulencia causada por constricciones en el tracto vocal (*no voceados*).

En el espectrograma de la figura 1.3 se pueden ver las diferencias espectrales cualitativas entre las dos categorías. La estructura espectral de los sonidos voceados consiste en bandas oscuras bien definidas, mientras que las consonantes (en el caso de la figura, ambas son indistinguibles para la palabra *voces* pronunciada en castellano) están caracterizadas por una mancha relativamente homogénea para frecuencias altas. La diferencia entre ambas se explicará en el capítulo 2, en el marco de la teoría de *fuentes-filtro*. Baste ahora con reconocer las diferencias entre las dos grandes clases de sonido presentadas.

La distinción lingüística de los sonidos de la lengua (*fonemas*) no es unánime. En casi todas las lenguas hay discrepancias no solo respecto del número de fonemas que las conforman, sino también de su clasificación. En este trabajo no nos ocupamos de estas entidades, que no siempre tienen correlato con las características puramente acústicas de los sonidos, aunque una descripción muy somera permitirá entender después la construcción de una base de datos de voz en la validación de los modelos desarrollados en los próximos capítulos.

Las *vocales* son los sonidos voceados por excelencia. Durante su pronunciación, el tracto nasal está acústicamente desacoplado del tracto vocal, que mantiene una configuración fija mientras es excitado por el flujo de aire variable

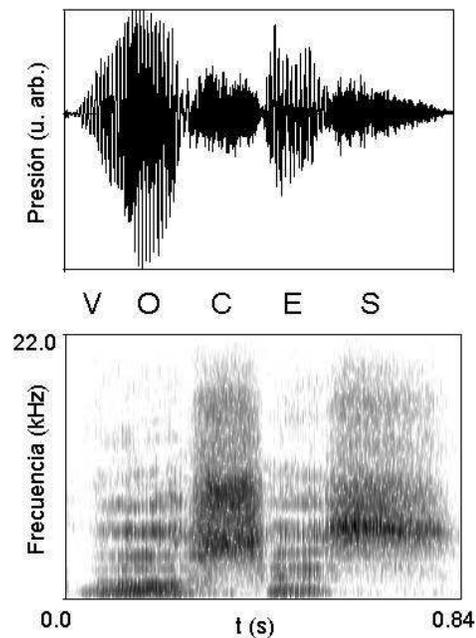


Figura 1.3: Serie temporal (envolvente) y sonograma correspondiente al registro de la palabra *voces*. La serie temporal fue registrada a una tasa de muestreo de 44100 Hz, y 16 bits de resolución. El sonograma está resuelto usando ventanas de 512 puntos, hasta una frecuencia máxima de 22050 Hz, la frecuencia de Nyquist. Durante la pronunciación de las vocales, el contenido espectral está caracterizado por bandas bien definidas de frecuencia, mientras que para las consonantes, en este caso, pueden ser caracterizadas espectralmente como ruido de alta frecuencia.

provocado por la vibración de las cuerdas vocales. Su duración durante el habla, relativa a la mayoría de las consonantes, es mucho mayor.

Otros sonidos voceados son los *diptongos*, durante cuya pronunciación el tracto vocal cambia desde una a otra vocal, como $/a^w/$ en ‘auto’, las *semivocales*, como la $/l/$ en ‘lengua’, caracterizadas por una configuración de transición entre fonemas adyacentes y las *consonantes nasales*, como la $/m/$ y la $/n/$ (‘mono’), para las que se usa el tracto nasal, aunque la cavidad oral está acústicamente acoplada y funciona como una caja de resonancia.

Los sonidos no voceados pueden clasificarse a su vez en varias categorías. Las *africativas no voceadas*, como la $/f/$ o la $/s/$, que generan sonido mediante turbulencias generadas en alguna constricción del tracto vocal, las *fricativas voceadas*, como la $/v/$, en las que participan las cuerdas vocales y también la turbulencia como fuentes de sonido y son de muy corta duración durante el discurso, y las *oclusivas*, como $/b/$ o $/p/$, son algunos ejemplos.

Las características espectrales bien definidas de los distintos sonidos del habla invitaron, a lo largo del tiempo, al desarrollo de distintas técnicas para tratar los espectros de las señales acústicas de la voz. En la sección siguiente se describe la matemática de algunas técnicas espectrales más usuales.

1.2. Técnicas de reconstrucción espectral

Supongamos que la serie temporal $h(t)$ es muestreada a intervalos temporales Δ . La tasa de muestreo ($1/\Delta$) define una frecuencia crítica, $f_c = 1/(2\Delta)$, tal que para funciones limitadas a frecuencias $|f| < f_c$, la función $h(t)$ puede ser reconstruida a partir de sus muestras. Este es el *teorema de muestreo*.

Este teorema impone un intervalo en el que las funciones espectrales discretas son confiables: el menor muestreo posible para una función armónica es de dos puntos por ciclo.

A lo largo de esta tesis nos enfocaremos al tratamiento de las series temporales de la voz en condiciones de tecnología de uso comercial. En este campo, la tecnología estándar es suficiente para el análisis de la voz; las computadoras personales equipadas con micrófonos comerciales trabajan con respuestas conocidas para el rango de frecuencias típico, y las tasas de muestreo garantizan resoluciones de hasta $f_c = 1/2\Delta = 22.050$ Hz.

En las próximos apartados revisaremos las técnicas más divulgadas de reconstrucción espectral con vistas a una introducción de los métodos de reconocimiento de la identidad por la voz, y a nuestro propio tratamiento del problema.

Transformada rápida de Fourier (FFT)

Esta es la técnica más usual para el tratamiento espectral de series temporales discretas $x_n = x(n\Delta)$. La transformada de Fourier discreta mapea la serie de N números x_n en otra de N números complejos X_n que se aproxima a la transformada de Fourier $X(f_n)$ evaluada en las frecuencias $f_n = n/(N\Delta)$ ($-N/2 \leq n \leq N/2$), es decir

$$X(f_n) = \int_{-\infty}^{\infty} x(t)e^{2\pi i f_n t} dt \approx \Delta \sum_{k=0}^{N-1} h_k e^{2\pi i k n / N} = \Delta X_n. \quad (1.2)$$

Existe una gran variedad de algoritmos para calcular la transformada de Fourier discreta [NR99]. Sin embargo, existen otras técnicas espectrales que permiten la construcción de funciones analíticas, computacionalmente económicas y rápidas. El ejemplo más extendido es la reconstrucción espectral a través de coeficientes de predicción lineal, o lpc.

Código de predicción lineal (lpc)

El filtro lineal más genérico toma una serie de N datos x_k y produce una secuencia y_n de salida según la fórmula

$$y_n = \sum_{k=0}^{N-1} c_k x_{n-k} + \sum_{j=0}^{M-1} d_j y_{n-j-1}. \quad (1.3)$$

Los coeficientes c_k se definen en términos de los N valores previos de entrada y los coeficientes d_j dependen de los M valores previos de salida. Estos coeficientes definen la respuesta del filtro. La relación entre los coeficiente c_k y d_j y la respuesta del filtro en frecuencias $H(f)$ es

$$H(f) = \frac{\sum_{k=0}^{N-1} c_k e^{-2\pi i k (f\Delta)}}{1 - \sum_{j=0}^{M-1} d_j e^{-2\pi i (j+1) (f\Delta)}} \quad (1.4)$$

Si todos los coeficientes d_i son cero, el numerador de (1.4) se reduce a la transformada de Fourier discreta. Inversamente, el código de predicción lineal busca el conjunto de coeficientes d_j (ec. 1.3) que minimizan x_n en

$$y_n = \sum_{k=1}^m d_k y_{n-k} + x_n. \quad (1.5)$$

de manera que la respuesta del filtro resulte una función algebraica. Los coeficientes d_i se calculan usando aproximaciones de la autocorrelación

$$\phi_j \equiv \langle y_i y_{i+j} \rangle \approx \frac{1}{N-j} \sum_{i=0}^{N-j-1} y_i y_{i+j} \quad (1.6)$$

y M ecuaciones con los coeficientes d_i como incógnitas,

$$\sum_{j=0}^{M-1} \phi_{|j-k-1|} d_j = \phi_k. \quad (1.7)$$

La función resultante puede extenderse a los complejos

$$H(z) = \frac{a_0}{1 - \sum_{j=1}^{M-1} d_j z^{j+1}} \quad (1.8)$$

con a_0 la amplitud media cuadrática de x . La función presenta M polos y periódica en el círculo unidad $z \equiv e^{2\pi i f \Delta}$, (figura 1.4).

La aproximación al espectro de potencias $|H(f)|^2$ calculada mediante los coeficientes *lpc* presentan varias ventajas respecto a otras aproximaciones. En general, la aproximación al espectro por funciones algebraicas (que presentan polos cuando se extienden a los complejos) permite capturar características agudas del espectro (por ejemplo, sus formantes) con pocos coeficientes. En los sistemas de reconocimiento de voz que estudiaremos en la próxima sección, esta economía es deseable para caracterizar espectralmente largas vocalizaciones.

En este trabajo, en cambio, estamos interesados en esta aproximación porque permite construir funciones periódicas, continuas y derivables que aproximan las características espectrales de la voz. Estas funciones permiten la construcción de curvas tridimensionales cerradas, cuya caracterización topológica exploraremos en el capítulo 4.

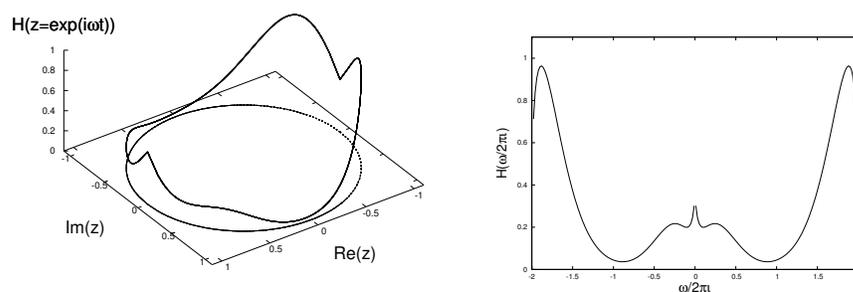


Figura 1.4: La función $H(z) \in \mathbb{C}$ definida en $z = e^{i\omega t}$ es periódica en el círculo unidad. En el panel izquierdo, $H(z)$ en el plano complejo, y en el panel derecho su representación en términos de la frecuencia ω .

Distancias Cepstrales

Es usual encontrar otras codificaciones para las propiedades espectrales de las series temporales de la voz. El *cepstrum* complejo de una señal, por ejemplo, está definido como la transformada de Fourier del logaritmo del espectro de potencias.

$$\log H(\omega) = \sum_{-\infty}^{\infty} c_n e^{-\omega n} \quad (1.9)$$

con $c_n = c_{-n} \in \mathbb{R}$, llamados coeficientes *cepstrales*.

Notablemente, se cumple que la distancia cuadrática d entre un dado par de espectros H y H' cumple

$$d(H, H') = \int_{-\pi}^{\pi} |\log(H(\omega)) - \log(H'(\omega))|^2 d\omega/2\pi = \sum_{-\infty}^{\infty} (c_n - c'_n)^2, \quad (1.10)$$

donde c_n y c'_n son los coeficientes cepstrales de $H(\omega)$ y $H'(\omega)$ respectivamente.

En particular, para modelos LPC que representan funciones espectrales suaves, la suma de la ecuación 1.10 puede ser truncada a un número pequeño (~ 10) de términos sin perder información.

Las estrategias presentadas aquí son sólo algunos ejemplos de una gran familia de caracterizaciones espectrales desarrolladas para tratar espectros de series de voz. En general, el espíritu detrás de todas ellas es la búsqueda de métodos computacionalmente económicos de comparación y almacenamiento de espectros.

En la siguiente sección revisaremos algunas generalidades del campo de la *biometría*, donde estas propiedades cobran sentido al formar parte de sistemas de reconocimiento por la voz.

1.3. Biometría

La biometría agrupa las áreas técnicas, científicas y tecnológicas que estudian propiedades biológicas y de comportamiento para la identificación de las personas. La firma, las huellas digitales y el análisis genético del ADN demuestran que se trata de un conjunto heterogéneo reunido por cierta consistencia en la repetitividad que permite su uso como validación de la identidad.

En términos generales, los métodos más seguros son invasivos y requieren unas condiciones específicas para la medición. El análisis de ADN y el scanning de retina son dos ilustraciones de este caso y coinciden, además, con la necesidad de equipamiento costoso. Por otra parte, la firma y las huellas digitales son características biométricas ampliamente usadas, pero su información resulta relativamente fácil de falsear.

La biométrica de la voz está en alguna parte entre esos casos extremos. En parte, esto se debe a que las mediciones -los registros de voz- no son tratados como provenientes de un aparato vocal, sino analizados y comparados por métodos estadísticos. En general, la biometría está inspirada en la propia percepción: nosotros reconocemos a las personas por su rostro, su voz, su forma de caminar. En el caso de la voz, los mecanismos de percepción dependen de factores de entonación y de colocación, pero además está ligada al conjunto de percepciones, y también a la memoria y a otras operaciones psíquicas. Existen, inspirados en los mecanismos auditivos y de procesamiento neuronal, métodos biométricos de tratamiento espectral para el reconocimiento de la voz. En el próximo punto analizaremos muy brevemente los avances históricos en el campo, antes de introducir la matemática del problema.

Las técnicas de reconocimiento de voz se han desarrollado por al menos cuatro décadas, y aun hoy es un área fecunda de trabajo. Los primeros intentos, en los albores del cincuenta, estaban apoyados en el reconocimiento de las resonancias espectrales durante la pronunciación de las vocales para un vocabulario seleccionado y bastante limitado. Los primeros avances se constituyeron sobre extensiones de los vocabularios, incorporando información estadística sobre las secuencias de fonemas usados.

Durante la década siguiente, los avances se focalizaron en técnicas marginales al problema que permitieran lidiar con escalas temporales no uniformes a lo largo de la serie temporal de voz y con el reconocimiento de fonemas en series de habla continua.

En la década del setenta, el desarrollo de técnicas espectrales como el código de predicción lineal (desarrollado en la sección 1.2) permitió el uso de medidas de distancia basadas en pocos parámetros espectrales. Además, se desarrollaron los sistemas de reconocimiento *independientes de texto* a través sofisticados algoritmos de *clustering* para determinar todas las variaciones de palabras a lo largo de una amplia base de usuarios.

En los ochenta, los métodos puramente estadísticos dieron paso al uso de algoritmos más complejos, como los modelos de Markov ocultos (HMM) y redes neuronales. De allí en adelante, los esfuerzos se consumieron en el desarrollo de

grandes bases de vocabulario y sistemas de reconocimiento de habla continua, junto con la exportación de los sistemas a aplicaciones masivas, como la telefonía.

1.3.1. Principios de reconocimiento de identidad por la VOZ

Clasificación de la tecnología

El reconocimiento de la identidad por la voz puede ser separado en métodos de *identificación* y de *verificación* del hablante. La identificación del hablante es el proceso por el cual se determina de cuál de los usuarios registrados proviene una dada enunciación. La verificación, en cambio, es el proceso de aceptación o rechazo de la identidad reclamada por una persona.

La diferencia fundamental entre ellos es la cantidad de alternativas para la decisión. Durante la identificación, el número de decisiones es igual a la población registrada, mientras que en la verificación hay solamente dos alternativas: aceptación o rechazo.

Los métodos de reconocimiento pueden dividirse también en métodos *independientes del texto* y *dependientes del texto*.

Hay algunas aplicaciones, como las forenses, en las que no pueden usarse textos predeterminados. Más aún, los hombres tenemos la capacidad de reconocer personas independientemente del contenido del habla. Por lo tanto, los métodos independientes del texto son objeto de grandes esfuerzos de investigación a los que nos sumamos en esta tesis, con aproximaciones *independientes del texto* basadas en la física de la voz y en la topología de los espectros.

Todas las categorías de los sistemas tradicionales de reconocimiento comparten una estructura básica de comparación entre propiedades espectrales, que se estudia a continuación.

Estructura básica de los sistemas de reconocimiento

En todos los sistemas de reconocimiento, los registros de voz de una persona cuya identidad se desconoce son analizados y comparados contra registros conocidos (bancos de voz). Si la comparación se encuentra por arriba de un cierto *umbral*, la identidad resulta verificada. Umbrales altos tienden a disminuir el riesgo de aceptar una identidad falsa (falso positivo) al costo de aumentar las probabilidades de rechazar a un usuario registrado (falso negativo).

La primera etapa del análisis consiste en la extracción de propiedades espectrales de la señal de habla original. Para eso se definen los *vectores característicos*, arreglos $\vec{d} = (d_1, \dots, d_n)$ de coeficientes espectrales d_j para cada ventana de análisis a lo largo de la serie temporal. Con esta operación se convierte una serie temporal del habla en una secuencia de vectores característicos para cada una de las ventanas de análisis.

Los coeficientes más utilizados para la creación de vectores característicos son los lpc y los coeficientes cepstrales. Dados dos vectores característicos \vec{x} y \vec{y}

definidos en un espacio vectorial χ , una *métrica* o *distancia* d en el espacio χ es una función real definida en $\chi \times \chi$ que cumple

- (a) $0 < d(\bar{x}, \bar{y}) < \infty$ para $\bar{x}, \bar{y} \in \chi$ y $d(\bar{x}, \bar{y}) = 0 \iff \bar{x} = \bar{y}$;
- (b) $d(\bar{x}, \bar{y}) = d(\bar{y}, \bar{x})$ para $\bar{x}, \bar{y} \in \chi$;
- (c) $d(\bar{x}, \bar{y}) \leq d(\bar{x}, \bar{z}) + d(\bar{y}, \bar{z})$ para $\bar{x}, \bar{y} \in \chi$;

Si una distancia cumple con estas propiedades se acostumbra a llamarla una *medida de distorsión*, particularmente cuando los vectores son representaciones del espectro de la voz. Para ser útil al procesamiento del habla, una medida de distancia debe tener algún correlato con juicios subjetivos que vuelven más flexibles las propiedades de las métricas [Tit93].

Las diferencias entre patrones espectrales del habla en términos de promedios o distorsión espectral acumulada aparece como una forma matemáticamente tratable y computacionalmente económica. Entre las medidas más usuales se cuentan las *distancias espectrales logarítmicas*, definidas por

$$d^p(H, H') = \left[\int_{-\pi}^{\pi} |\log(H(\omega)) - \log(H'(\omega))|^p d\omega/2\pi \right]^{1/p} \quad (1.11)$$

para $p \in N$, y las *medidas cepstrales*

$$d_{2C}^2(H, H') = \int_{-\pi}^{\pi} \left| \frac{d \log(H(\omega))}{d\omega} - \frac{d \log(H'(\omega))}{d\omega} \right|^2 d\omega/2\pi \quad (1.12)$$

basadas en las diferencias entre pendientes espectrales. Esta distancia también se puede escribir a partir de los coeficientes cepstrales c_n

$$d_{2C}^2(H, H') = \sum_{-\infty}^{\infty} n^2 (c_n - c'_n)^2. \quad (1.13)$$

Esquema genérico de los sistemas de reconocimiento

El funcionamiento de los sistemas de reconocimiento puede describirse en las siguientes etapas:

- A partir de una enunciación nueva en el sistema, se descompone la señal temporal en ventanas de análisis del orden de las decenas de milisegundos.
- Para cada ventana se calcula un vector característico, usando coeficientes lpc o derivados de lpc (como Cepstrum).
- Se construyen templados a partir de los vectores característicos.
- Se comparan los vectores característicos calculados con los archivados en el banco de datos. Los templados almacenados se construyen promediando los templados de varias sesiones de entrada para cada usuario (entrenamiento).

- Se establece la distancia entre el templado de entrada y los del banco a través de alguna de las medidas de distorsión.
- Se comparan las distancias contra el umbral de aceptación, que define si la entrada es aceptada como uno de los registros previos, o rechazada.

Este es un cuadro general del proceso de identificación. Para el caso de sistemas de verificación del hablante, sólo cambia la comparación, que se ejecuta contra el usuario cuya identidad se desea validar.

Del resumen presentado en este capítulo surge inmediatamente la sensación de una profunda desvinculación entre la fisiología de la voz humana y las técnicas usadas para su uso biométrico. Casi toda la historia de la biometría consiste en la producción de grandes familias de herramientas espectrales cuyo acercamiento a la fisiología, si alguno, está puesto en el correlato perceptual. En este sentido, han sido propuestas medidas de distorsión psicofísicas, o sintonizadas a las respuestas en frecuencia del oído (los coeficientes MEL-cepstrum se construyen usando una curva de respuesta como la que caracteriza el oído [Rab93]).

En el próximo capítulo trataremos un enfoque distinto, para el que construiremos sobre esfuerzos previos ordenados a la descripción de la voz en términos de la física del aparato vocal. Construiremos un modelo completo para la producción de la voz con parámetros que pueden correlacionarse con las magnitudes fisiológicas del problema, y propondremos un método de reconocimiento de la identidad basado el ajuste de estos parámetros a partir de registros de voz.

Capítulo 2

Física de la producción de la voz

En este capítulo se presenta un modelo no lineal minimal para la dinámica de las cuerdas vocales, controlado por parámetros que tienen un correlato con la fisiología descrita en el capítulo anterior.

Los sistemas no lineales pueden mostrar comportamientos cualitativamente distintos al variar sus parámetros, por lo que su análisis permite acotar los tipos de dinámica admitidas. Incorporamos luego el efecto del tracto vocal en la señal generada por la oscilación de las cuerdas vocales, lo que permite generar señales vocales sintéticas de voz. Los resultados de este capítulo fueron compilados en [TEM01].

Interacción entre laringe y tracto vocal

El aparato vocal descrito en el capítulo 1 admite ciertas analogías con algunos instrumentos de viento. Sin embargo, mientras la mayoría esos instrumentos consiste en osciladores fuertemente acoplados al tubo resonador, el aparato fonador admite distintos grados de acoplamiento entre las cuerdas y el tracto vocal [Tit93]. En el caso del habla el acoplamiento es débil [LajMin01], y las cuerdas vocales crean sus propias condiciones de oscilación a través de sus modos normales. La onda de presión generada por el flujo de aire a través de la glotis viaja por el tracto vocal, cuya geometría es modificada por sus articuladores principales. Los armónicos enfatizados reciben el nombre de *formantes*.

A través de los articuladores, el tracto configura distintas geometrías para las cuales la laringe puede mantener y controlar su frecuencia fundamental e intensidad sin interferencias del tracto vocal. Este desacople, que favorece la constancia en la fonación, es la base de la teoría de *fuentes-filtro*, que inspira los modelos del aparato fonador y que exploraremos a continuación.

2.1. Modelo de Fuente-Filtro

2.1.1. Fuente: modelos de cuerdas vocales

La construcción de modelos vocales tiene una historia muy rica. En todos ellos, el flujo de aire induce inestabilidades en las cuerdas vocales, creando un flujo variable que excita al tracto vocal. En un trabajo fundacional, Ishizaka y Flanagan [IshFla72] construyeron un modelo de dos masas acopladas, que sigue siendo usado tres décadas después de su publicación. Manteniendo la simplicidad del problema inicial, Story y Titze introdujeron un sistema de tres masas que permitió una mejor correspondencia entre los parámetros del modelo y las variables fisiológicas involucradas [StoTit95].

En esta sección reseñaremos algunos de los modelos existentes para enmarcar los mecanismos principales del inicio y autosustentación de las oscilaciones en las cuerdas vocales. Todos los modelos asumen el principio de oscilación inducida por flujo de aire, sin clausura glotal completa.

El modelo más simple es conocido como *modelo de una masa*. Cada cuerda vocal corresponde a una masa ligada elásticamente al tejido laríngeo (ver sección 1.1.1). En este modelo es posible mostrar que la presión sobre las cuerdas vocales (intraglotal) es la misma que a la entrada del tracto vocal (figura 2.2). Para generar inestabilidades y transferir energía desde el flujo de aire al tejido es necesario que durante el semiciclo de apertura de las cuerdas la presión sea mayor que durante el cierre. Estas condiciones se pueden alcanzar siempre que el aire en el tracto vocal sea principalmente inercial. Este modelo será desarrollado en la próxima sección.

También es posible generar oscilaciones independientes del acople con el tracto vocal. La forma más simple de hacerlo en ausencia de acople es asumir una estructura de cuerdas vocales no uniforme [Tit88].

La idea consiste en que la forma de la glotis puede cambiar en un ciclo de oscilación (figura 2.1). En este esquema, la glotis presenta un perfil convergente al abrirse, y la presión intraglotal será parecida a la presión subglotal durante este semiciclo. Al tomar un perfil divergente durante el cierre, la presión será más parecida a la atmosférica. En el ciclo completo, la glotis recibe una fuerza neta mayor en la dirección del movimiento durante la apertura, y se transferirá energía desde el flujo de aire al tejido, sosteniendo las oscilaciones.

Sistema dinámico del modelo de una masa

En este trabajo nos concentramos en el modelo más simple y mostraremos que las oscilaciones no lineales que resultan son creadas en una bifurcación de Hopf, y pueden generar, una vez inyectadas en el tracto vocal, series sintéticas de presión como las registradas por un micrófono.

Cada cuerda vocal se asume como una masa m sujeta al tejido principal a través de una fuerza restitutiva elástica, una fuerza disipativa, y la fuerza ejercida por la presión intraglotal. El modelo se esquematiza en la figura 2.2.

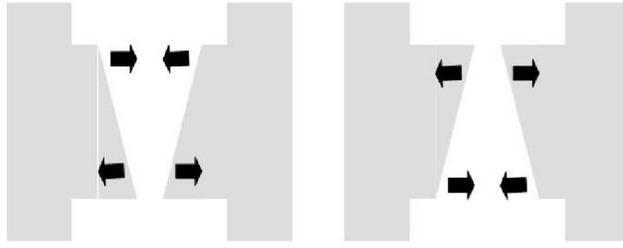


Figura 2.1: En el modelo de dos masas el perfil de las cuerdas vocales, convergente durante el semiciclo de apertura y divergente durante la clausura asegura la transferencia de energía desde el flujo de aire necesaria para sostener las oscilaciones.

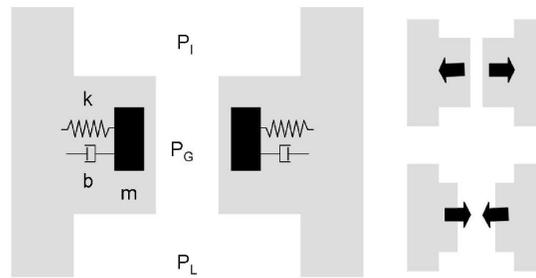


Figura 2.2: Esquema del modelo de una masa para la dinámica de las cuerdas vocales. El sistema es simétrico y consiste de masas m sostenidas al tejido principal a través de una fuerza restitutiva de constante k . Los elementos anatómicos de los que se extrae esta caricatura están graficados en la figura 1.1.

En este modelo se puede mostrar que la presión intraglotal es la misma que a la entrada del tracto vocal. Para generar el comportamiento oscilatorio, es necesario que la presión durante la apertura de las cuerdas vocales sea mayor que la atmosférica, y menor en el semiciclo de cierre. Este modelo constituye una buena aproximación mientras que el aire en el tracto tenga propiedades inertivas. Cuando las cuerdas se están separando, el flujo aumenta y la columna de aire en el tracto se acelera, lo cual tiende a abrir más aun las cuerdas. En el semiciclo de cierre, por otra parte, la columna de aire en movimiento tiende a cerrar más aun las cuerdas.

Las ecuaciones de movimiento para este modelo de cuerdas vocales pueden ser escritas como

$$m\ddot{x} + b\dot{x} + kx = p_g \quad (2.1)$$

donde m , b y k representan la masa, la constante de disipación y la constante

de restitución, por unidad de área, de las cuerdas vocales, y p_g la presión glotal. La variable x describe el apartamiento de las cuerdas vocales de su posición de equilibrio.

En el caso general para el que el área de entrada a la glotis (a_1) y de salida (a_2) son distintas, se mostró [Tit93] que la relación entre la presión transglotal y la intraglotal es

$$p_g = p_i + (p_s - p_i)(1 - a_2/a_1 - k_e)/k_t, \quad (2.2)$$

con p_i la presión a la entrada del tracto vocal y p_s la presión subglotal. Los coeficientes k_e y k_t son coeficientes fenomenológicos que dan cuenta de las diferencias entre presión y velocidad esperadas para flujos estacionarios y se conocen como *coeficiente de recuperación pulmonar* y *coeficiente de presión transglotal* respectivamente.

Dado que en el modelo de una masa las áreas de entrada y salida al tracto vocal son iguales ($a_1 = a_2$) y $k_e \sim 0$ cuando el área glotal es mucho más pequeña que la del tracto vocal ($a_2 \ll S_t$) se verifica $p_i = p_g$.

El último elemento necesario para la construcción del modelo para la dinámica de las cuerdas vocales (x) es una relación entre la presión y las variables intraglotales. El elemento clave es que siempre que la frecuencia fundamental de oscilación de las cuerdas vocales sea menor que la primera resonancia del tracto (primera formante), su impedancia de entrada es principalmente inercial [Rot81], de manera que se puede escribir

$$p_i = R_2U + I_2U', \quad (2.3)$$

con R_2 e I_2 la resistencia y la constante de inercia del aire en el tracto, y U el flujo. El sistema dinámico queda cerrado, entonces, relacionando las ecuaciones 2.1 y 2.3 mediante una modificación de la ecuación de Bernoulli

$$p_L - (R_2U + I_2U') = k_t\rho v^2 \quad (2.4)$$

donde $k_t \sim 1$ es un factor de corrección fenomenológico y v la velocidad del aire en la glotis ($v = U/a$). Notamos aquí que el área glotal puede escribirse en términos de la longitud de las cuerdas vocales L y la posición de equilibrio de las cuerdas x_o como $a = 2L(x + x_o)$.

De la integración de las ecuaciones 2.1, 2.3 y 2.4, es posible escribir un sistema de tres ecuaciones de primer orden. Definiendo

$$F = p_L - \rho/2v^2 - 2L[R_2(x_o + x) + I_2y]v/[2L(x_o + x)I_2] \quad (2.5)$$

el sistema de ecuaciones para la dinámica de las cuerdas vocales en el modelo de una masa es:

$$\dot{x} = y, \quad (2.6)$$

$$\dot{y} = 1/m[2LR_2x_0v + 2LI_2x_0F - (b - 2LI_2v)y - (K - 2LR_2v - 2LI_2F)x], \quad (2.7)$$

$$\dot{v} = F. \quad (2.8)$$

Este sistema de ecuaciones fue derivado bajo la hipótesis de que las cuerdas vocales no colapsan (como queda implicado, por ejemplo, en la relación entre p_L y p_s usando una versión corregida de la ecuación de Bernoulli). Por tanto, esperamos soluciones relevantes solamente para $x > -x_0$.

Análisis del sistema dinámico

Para estudiar los distintos regímenes de soluciones del sistema, elegimos los parámetros de restitución k y presión pulmonar p_L , que representan las cantidades críticas para la fonación (sección 1.1.1).

En general el sistema tiene hasta tres puntos estacionarios, aunque solo dos en el dominio $x > -x_0$, uno para valores positivos de flujo, $v^* = v_1$, y otro para valores negativos, $v^* = v_2$.

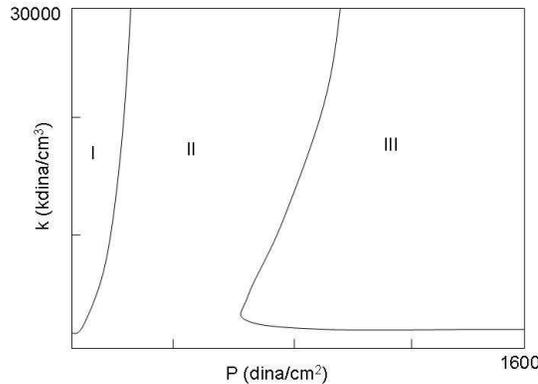


Figura 2.3: Diagrama de bifurcaciones del modelo de una masa. En la región I, la aparición de un ciclo límite atractor representa el inicio de oscilaciones sostenidas en las cuerdas vocales.

Comenzamos la descripción por la zona I (figura 2.3). En esta región, el punto fijo en $v_1 > 0$ es un nodo silla. Tiene una dirección estable (aproximadamente paralela al eje v), y una variedad inestable bidimensional (asociada con autovalores complejos conjugados) que alimentan a un ciclo límite. El punto coexistente v_2 es un saddle. Localmente, tiene una variedad estable aproximadamente paralela al eje x , y dos direcciones inestables (aproximadamente paralelas al eje y y v). Una variedad bidimensional, tangente en el punto fijo v_2 a su variedad

estable y a la primera dirección inestable descripta, separa el espacio de fases entre la cuenca de atracción del ciclo límite y los puntos que tienden al infinito. Esta variedad se aproxima al plano singular $x = -E_0$ en la curva en la que el numerador de F se anula.

En la figura 2.4 se ve una proyección bidimensional del flujo en (x, v) . Se grafica una parte de la cuenca de atracción junto con una trayectoria de condiciones iniciales justo por encima, aproximándose al ciclo límite. La línea que separa las regiones I y II indica la ocurrencia de una bifurcación de Hopf en la que el ciclo límite se crea desde $v_1 > 0$. En otras palabras, en la región II, el punto fijo $v_1 > 0$ es un atractor.

Por completud, terminamos la descripción con los cambios en la cuenca del infinito. Hacia la región III, el punto fijo $v_2 < 0$ tiene dos autovalores que se vuelven complejos conjugados (con autovectores asociados definiendo un plano casi paralelo del plano (\dot{x}, z)). En la separatriz entre las regiones II y III ocurre una bifurcación de Hopf inversa: el punto $v_2 < 0$ emite un ciclo límite saddle, ganando estabilidad.

En resumen, este modelo predice la aparición de oscilaciones sostenidas a medida que el flujo aumenta. Desde el punto de vista dinámico, la oscilación se crea en una bifurcación de Hopf y por lo tanto la solución, a medida que el flujo se incrementa, cambia de frecuencia y de contenido espectral [GucHol83] a medida que se establece una oscilación no lineal.

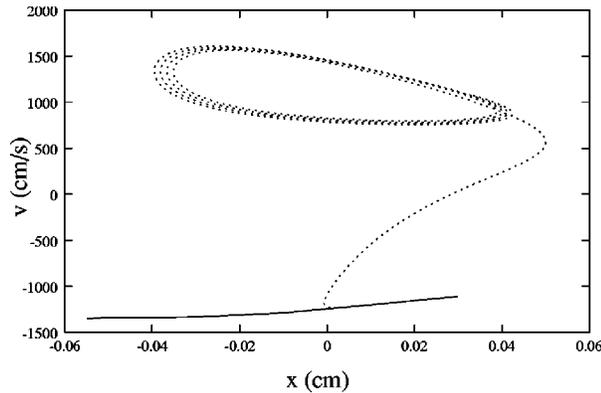


Figura 2.4: Una proyección bidimensional del flujo en el espacio (x, v) . Se muestra una parte de la cuenca de atracción junto con la trayectoria correspondiente a una condición inicial justo por encima de ella, acercándose al ciclo límite atractor. Los parámetros usados en la simulación son $L=0.14$ cm, $m=0.04$ g/cm², $k = 20 \cdot 10^3$ kdina/cm³, $x_0 = 0.06$ cm, $p_L = 840$ dina/cm², $\rho=0,00114$ g/cm³, $R_2=3.1$ dina/cm⁵, $I_2=0.1$ dina s²/cm⁵ y $c=35000$ cm/s.

2.1.2. Filtro: el tracto vocal

Siguiendo la teoría *fuentes-filtro*, el desacople acústico entre las cuerdas vocales y el tracto vocal permite generar una señal de voz a través del filtrado de la serie temporal que se genera por la variación de presión P_i a la entrada del tracto.

En su versión más simplificada, el tracto vocal puede pensarse como un tubo de sección uniforme, por donde viaja la señal de presión $P_i(t)$ originada en la oscilación de las cuerdas vocales a la entrada del tracto. En esta aproximación, uno de los extremos del tubo está cerrado (el área de la sección transversal de la laringe es despreciable frente a la de entrada del tracto), y el otro, a la salida de la boca, abierto.

Filtrado de la serie temporal

Esta aproximación elemental al tracto vocal por un único tubo de sección uniforme puede ser sistemáticamente mejorado para representar un tracto vocal de sección variable durante el habla. Las aproximaciones usuales del tracto vocal se componen de una serie de N tubos de distinta sección A_i y longitud L_i ($i = 1, \dots, N$) para simular el efecto de la lengua, la mandíbula y los labios. En cada interfaz, la onda de presión es parcialmente reflejada y transmitida. El coeficiente de reflexión para la primera interfaz, por ejemplo, está dado por $r_{1,2} = (A_1 - A_2)/(A_1 + A_2)$, y el de transmisión, $t_{1,2} = 1 - r_{1,2}$. Para un sistema de 3 tubos, las ecuaciones que describen el filtrado de la onda sonora son:

$$a(t) = P_i(t) + b_b(t - \tau_1), \quad (2.9)$$

$$b_b(t) = r_{1,2}a(t - \tau_1) + t_{2,1}c_b(t - \tau_2), \quad (2.10)$$

$$b_f(t) = t_{1,2}a(t - \tau_1) + r_{2,1}c_b(t - \tau_2), \quad (2.11)$$

$$c_b(t) = r_{2,3}b_f(t - \tau_2) + t_{3,2}d_b(t - \tau_3), \quad (2.12)$$

$$c_f(t) = t_{2,3}b_f(t - \tau_2) + r_{3,2}d_b(t - \tau_3), \quad (2.13)$$

$$d_b(t) = \alpha c_f(t - \tau_3), \quad (2.14)$$

con $\alpha \sim -1$ el coeficiente de reflexión entre la atmósfera y el tercer tubo, y $\tau_i = L_i/c$ el tiempo que tarda la onda sonora de velocidad $c \sim 340m/s$ en recorrer el tubo de longitud L_i .

En la figura 2.5 se muestran series temporales correspondientes a la presión (grabada por un micrófono) y la generada por el modelo (ecs. 2.14) al inicio de la fonación (alrededor de 5 ms). Para la señal sintética se usó $p_i(t)$ como señal de entrada según la ecuación 2.8, es decir, el modelo completo para el aparato vocal. En ambas series temporales se ve una oscilación inicial simple que crece en amplitud y desarrolla riqueza espectral conforme pasa el tiempo. En términos del modelo, este fenómeno puede explicarse dinámicamente por la pérdida de estabilidad de un punto fijo en una bifurcación de Hopf, enriqueciendo espectralmente la oscilación, de la que el tracto vocal selecciona los armónicos según su configuración.

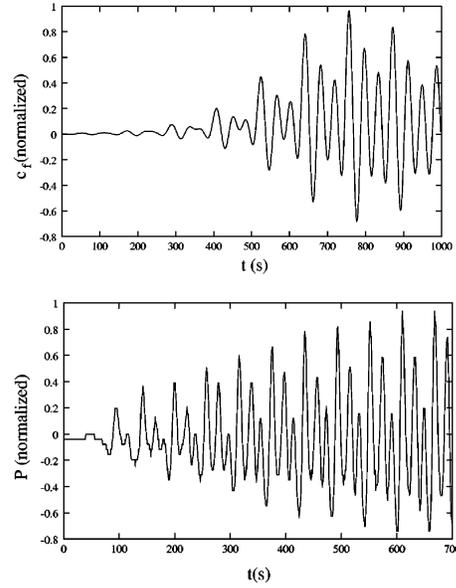


Figura 2.5: Registro experimental de la presión acústica al inicio de la fonación, correspondiente a la vocal /u/ (izquierda). Serie temporal numérica correspondiente a $c_f(t)$ (derecha).

Tracto vocal en el dominio de la frecuencia

En el dominio espectral, la aproximación del tracto por un tubo de sección uniforme corresponde a un tubo de cuarto de onda que presenta resonancias en $F_n = (2n - 1)c/4L$ Hz para $n = 1, 2, \dots, n$. Para un tubo de longitud $L = 17,5$ cm y velocidad del sonido $c = 350$ m/s, estas resonancias (o formantes) resultan $F_n = (2n - 1) \cdot 500$ Hz. Fuera de ellas, el espectro de frecuencias es cero. Sin embargo, aún en el caso idealizado, parte de la energía es irradiada por la boca, y parte también vuelve a los pulmones por la glotis. El tubo es, en consecuencia, menos selectivo en frecuencias, y el espectro resulta suavizado. Este espectro puede verse en la figura 4.3 con líneas punteadas.

A lo largo de este trabajo, pero especialmente en la aproximación topológica desarrollada en el capítulo 4, consideraremos el espectro de este tubo uniforme, no articulado, como un *espectro de referencia*.

El tratamiento espectral del filtrado de la presión a través de un tracto con una geometría general está relacionado con su evolución temporal a través de la convolución

$$c_f(t) = p_i(t) * h(t) \equiv \int_{-\infty}^{\infty} p_i(t - t')h(t')dt' \leftrightarrow C_f(\omega) = P_i(\omega) \cdot H(\omega), \quad (2.15)$$

donde usamos la notación anterior para la señal de entrada $p_i(t)$ y de salida

$c_f(t)$ del filtro. La respuesta temporal $h(t)$ del filtro representa las reflexiones y transmisiones a lo largo del tubo, mientras que su transformada de Fourier, $H(\omega)$, su respuesta espectral.

De este modo, en el dominio de las frecuencias, el filtrado puede verse como el énfasis o la supresión de las distintas componentes armónicas $P_i(\omega)$ de la señal de entrada de acuerdo a la respuesta espectral del tracto, $H(\omega)$.

En los modelos descriptos aquí se asume un desacople entre la fuente y el filtro. En este contexto, el sistema vocal está formado por una fuente no lineal activa de sonido (la glotis) y un filtro pasivo (el tracto vocal). Esta hipótesis es válida en condiciones normales de habla: el área glotal media es mucho menor que la sección de entrada al tracto vocal, y la frecuencia fundamental de oscilaciones es mucho menor que la primera formante o resonancia del tracto [Fla72]. En estas condiciones resulta indiferente trabajar en el dominio temporal o en el de la frecuencia para tratar las series temporales de la voz, y este último caso resulta particularmente útil porque el espectro de la voz se puede tratar como el producto entre el espectro de la fuente y el del filtro.

Este esquema no se puede aplicar cuando se tratan los efectos de retroalimentación del filtro sobre la fuente. En este caso, el único modelado posible es el de la integración de las ecuaciones diferenciales para la fuente incluyendo términos de la señal retrasada en el tiempo debido a las reflexiones en el tracto vocal [LajMin01]. En estas condiciones, el sistema presenta bifurcaciones que originan duplicaciones de período, subarmónicos y otros fenómenos que se han observado en la voz (por ejemplo, durante el canto armónico).

2.2. Ajuste de parámetros

Una vez construido el modelo completo para generar series temporales sintéticas de voz, naturalmente surge ponerlo a prueba como un método biométrico para reconocer la identidad de las personas, aprovechando que sus parámetros están relacionados con magnitudes anatómicas y fisiológicas del sistema de producción de la voz.

Sin embargo, la elección del método de ajuste es delicada porque interviene el conjunto de parámetros completo de la fuente y el filtro. La preferencia por un algoritmo genético se ve justificada por dos razones: es de programación simple y es fácilmente adaptable a la paralelización.

2.2.1. Algoritmos genéticos

Los algoritmos genéticos corresponden a la clase de métodos estocásticos de búsqueda. Mientras la mayoría de estos métodos operan sobre una única solución, estos algoritmos operan en una población de soluciones.

Crudamente, la idea biológica que los inspira es que el contenido genético de una población contiene potencialmente la solución, o una solución mejor que la de cada individuo particular, para la adaptación. Esta solución puede

estar inactiva porque la combinación genética adecuada está diseminada entre varios sujetos. Sólo la asociación de genomas distintos puede llevar a la solución [Mit98].

Aunque en términos biológicos el problema es mucho más complejo, una breve reseña es posible.

Los cromosomas son colecciones de genes que codifican las características del individuo. Cada gen, a su vez, es una secuencia de *bases nitrogenadas* que codifica una característica del individuo; su color de ojos, su altura, el tono de su piel.

Las bases nitrogenadas son cuatro: adenocina, citocina, timina y guanina. Con el alfabeto A, T, C y G está escrita toda la información genética en forma de genes que se agrupan en cromosomas.

A lo largo de la evolución los individuos se seleccionan, según su adaptación al ambiente, para la reproducción. Una característica general de la selección biológica es que, al tiempo que tiende a garantizar que la información genética de los mejores individuos se conserve a lo largo de las generaciones, también recombina la información distribuida en el resto de los individuos.

Una vez seleccionados los individuos, su información pasa a la siguiente generación modificada según dos operadores genéticos. El cruzamiento (o *crossing over*), consiste en el intercambio de información genética entre los individuos seleccionados para la reproducción. Las mutaciones, a su vez, introducen cierta aleatoriedad en la búsqueda cambiando uno de los genes por otro elegido al azar, contribuyendo a la exploración en el espacio de parámetros.

Las bases teóricas de funcionamiento y convergencia de los problemas de optimización tratados mediante algoritmos genéticos se discuten brevemente en el apéndice A.

En la siguiente sección se describe la adaptación de este algoritmo para el ajuste de parámetros en el problema de la voz.

2.2.2. Adaptación al problema de la producción de voz

El problema de la voz es fácilmente adaptable para ser resuelto por un algoritmo genético. En primer lugar, definimos un cromosoma como una cadena de números que corresponde a los parámetros del modelo de la voz. Así, la cadena $\{m, k, b, R_2, I_2, A_1/A_2, A_2/A_3, L_1, L_2\}$ es un cromosoma. m, k, b son los parámetros de las cuerdas vocales (ecuaciones 2.1 y 2.3); R_2, I_2, A_i y l_i los parámetros correspondientes al tracto vocal. En este trabajo elegimos aproximaciones del tracto por tres tubos restringidos a una longitud máxima $L_1 + L_2 + L_3 = 17,5$ cm.

El criterio de selección elegido consiste en considerar la probabilidad de seleccionar un cromosoma como inversamente proporcional a la diferencia cuadrática χ^2 entre la serie temporal que genera el modelo con los parámetros del cromosoma y la serie temporal experimental.

En la figura 2.6 se ve la evolución del mejor cromosoma de cada generación medido en términos de χ^2 para el primer segmento voceado de la figura 2.7. La

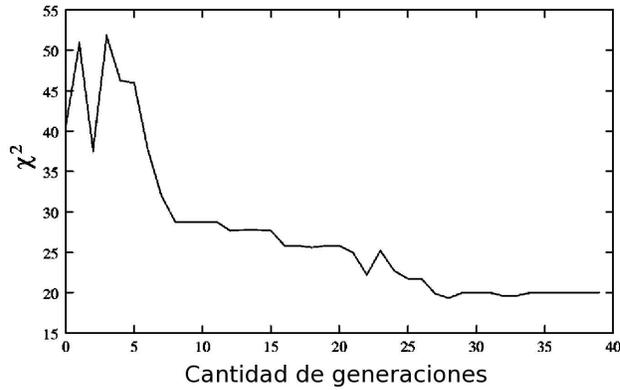


Figura 2.6: La evolución de χ^2 para el mejor cromosoma de cada generación. Se busca ajustar un segmento de aproximadamente tres oscilaciones correspondientes a un registro experimental de la vocal /e/ mostrado en la figura 2.7.

simulación del modelo con los parámetros del mejor cromosoma para la décima generación se ven en la figura 2.7 para distintos registros de sonidos vocados.

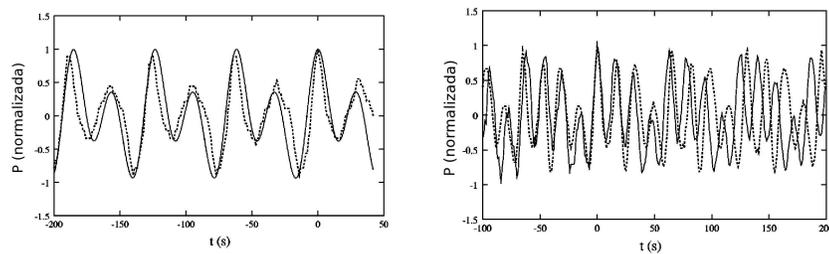


Figura 2.7: Dos segmentos vocados correspondientes a vocales del mismo sujeto y sus ajustes usando el algoritmo genético.

Para las vocales que muestran formantes muy distantes entre sí, la distancia cuadrática entre las series temporales no es la mejor medida para el ajuste de parámetros. Se ensayaron entonces ajustes entre los espectros de las series experimentales y sintéticas. En la figura 2.8 se muestra un ejemplo de ajuste para el espectro de una vocal /o/.

2.3. Resultados

Los resultados fundamentales del desarrollo del modelo presentado en este capítulo pueden resumirse en los siguientes puntos.

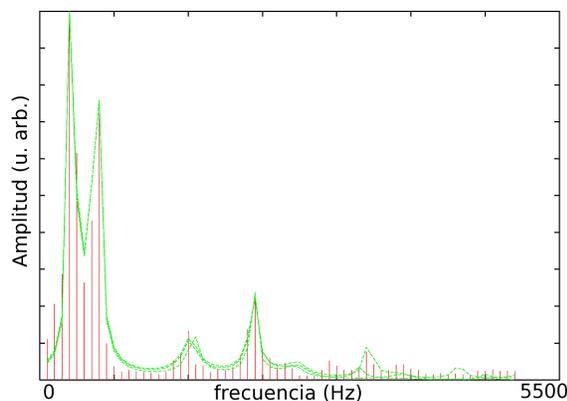


Figura 2.8: Espectro experimental de una vocal /o/ (barras verticales) y de los mejores espectros sintéticos (líneas), usando 500 generaciones y 150 cromosomas, tasa de mutación de 8% y crossover de 40%. Los espectros corresponden a soluciones con parámetros de la fuente distintos y distancias cuadráticas similares.

- La dinámica de las cuerdas vocales puede ponerse en términos del diagrama de bifurcaciones del modelo propuesto. Las inestabilidades causadas por el flujo de aire provocan las oscilaciones autosostenidas de las cuerdas vocales que se asocian a una bifurcación de Hopf del sistema dinámico. La presión crítica para el inicio de fonación crece con la tensión de las cuerdas vocales.
- Se puede reproducir el ‘ataque’ o inicio de los sonidos voceados. La serie resulta de los efectos del enriquecimiento espectral de la onda sonora a medida que es filtrada por el tracto vocal.
- Los algoritmos genéticos se muestran particularmente útiles para ajustar los parámetros del modelo a registros experimentales de sonidos voceados. Resultan fácilmente paralelizables (ver apéndice A.1), y el uso de poblaciones completas permite identificar los parámetros robustos del modelo.
- El modelo del aparato vocal muestra una mayor sensibilidad a la geometría del tracto vocal que a los de la anatomía de las cuerdas vocales 2.8. Los parámetros de las cuerdas vocales ajustados por el algoritmo presentan superposiciones para distintos hablantes mientras que los parámetros geométricos del tracto, además de corresponder a las configuraciones citadas por la bibliografía para cada vocal [Rab93], presentan una correlación más fuerte con la identidad del hablante, agrupándose en conjuntos mejor definidos para cada hablante.

En este capítulo construimos un sistema biométrico que permite un correlato más directo con la fisiología de la voz que los métodos espectrales desarrollados en el paradigma actual. El modelo usado captura los fundamentos de la dinámica de las cuerdas vocales y del filtrado por el tracto vocal usando una cantidad mínima de parámetros de control.

Sometiendo al modelo completo a la comparación con registros experimentales, los resultados sugieren que las características del tracto vocal son mejores indicadores biométricos de la identidad que los de la fuente de sonido. Estas observaciones invitan a una nueva interpretación del problema de la identificación de la identidad por la voz, que será presentado en el capítulo 4, después de describir los elementos necesarios en el próximo capítulo.

Capítulo 3

Sistemas dinámicos y topología

*‘trenzas,
nudo atroz de cuero crudo
que me ataron a tu mudo adiós...’*

Trenzas

Pontier, Expósito, 1945.

La aproximación espectral que propondremos en este trabajo está apoyada en un programa de caracterización las órbitas de sistemas dinámicos. Esta caracterización es de naturaleza topológica, y permite una descripción algebraica de la estructura en que las órbitas se conectan o anudan unas a otras. En este capítulo presentamos algunas herramientas de ese programa aplicadas a los sistemas dinámicos, que adaptaremos en el próximo capítulo al problema de la voz.

La última parte de esta tesis se dedica a una aproximación al control neuronal vocal. Para construir los modelos neuronales presentamos también las características de las soluciones de sistemas dinámicos forzados de baja dimensión.

3.1. Sistemas dinámicos forzados

Sometidos a forzados periódicos, tanto los *osciladores no lineales* como los *sistemas excitables* muestran una variedad de respuestas cualitativamente distintas, desde estructuras subarmónicas hasta la aparición de fenómenos caóticos.

Afortunadamente, la organización de esas respuestas presenta propiedades universales que permiten descripciones de naturaleza cualitativa y cuantitativa para todos los sistemas no lineales forzados. El objeto de esta sección es presentar algunas de estas propiedades que se usarán para modelar los mecanismos de control neuronal vocal en el capítulo 5.

Osciladores lineales

El forzado de un oscilador lineal por una función armónica de frecuencia ω_f esta descrito por

$$\ddot{x} + \gamma\dot{x} + \omega_o^2 x = A \cos(\omega_f t). \quad (3.1)$$

Después del transitorio, la solución resulta también una función armónica cuya frecuencia es la misma que la del forzado ω_f , es decir, $x_{part}(t) = A' \cos(\omega_f t + \phi)$. Esta solución está caracterizada por una amplitud A' y una fase ϕ dependientes de la frecuencia según

$$A'(\omega) = \frac{1}{\sqrt{(\omega_o^2 - \omega^2)^2 + \gamma\omega^2}} \quad \phi(\omega) = \frac{\gamma\omega}{\omega_o^2 - \omega^2} \quad (3.2)$$

de manera que, para un forzado periódico general $f(t) = a_o + \sum_{n=1}^{\infty} (a_n \cos(n\omega t) + b_n \sin(n\omega t))$, los sistemas lineales responden a la misma frecuencia fundamental, enfatizando o suprimiendo sus componentes armónicas, análogamente al filtrado de la señal sonora por el tracto vocal tratado en la sección 2.1.2.

Osciladores no lineales

Para osciladores no lineales, las respuestas posibles son mucho más variadas y han sido ampliamente estudiadas [SNM96, GucHol83, FGPV88], presentando fenómenos de caos, duplicaciones de período y soluciones subarmónicas.

Al forzar un oscilador no lineal, la solución más simple es una en que el oscilador y el forzante están en el caso equivalente al de los sistemas lineales, cuando las soluciones del sistema se repiten a sí mismas en el período del forzante.

Sin embargo, las soluciones pueden presentar casos de ‘enganches’ más complicados entre el sistema y su forzante. Estas soluciones subarmónicas estables se indican con los números de rotación $r = a/b$ para las soluciones que presentan a ciclos del forzado (en unidades del forzante) por cada b períodos del forzante. En la figura 3.1 se grafican las soluciones en el espacio (ω_f, A) para el oscilador de Van der Pol forzado. Las regiones bien definidas que aparecen se llaman *lenguas de Arnold* [GucHol83]. Para amplitudes A mayores, las lenguas se deforman y entrecruzan de distinta forma para cada oscilador particular.

Sistemas excitables

Los sistemas excitables se caracterizan por su comportamiento frente a estímulos externos [FGPV88]: cuando un sistema excitable es estimulado, existe una amplitud que separa dos tipos de respuestas. Para estímulos de tipo *subumbral*, las el sistema se mantiene cerca de su estado natural. Para estímulos de tipo *supraumbral*, en cambio, el sistema reacciona con una excursión en el espacio de fases antes de regresar a su estado inicial.

Cuando son forzados periódicamente, los sistemas excitables también presentan fenómenos de cuasiperiodicidad, duplicación de período, caos, y subarmónica, aún cuando el sistema no presenta una frecuencia natural.

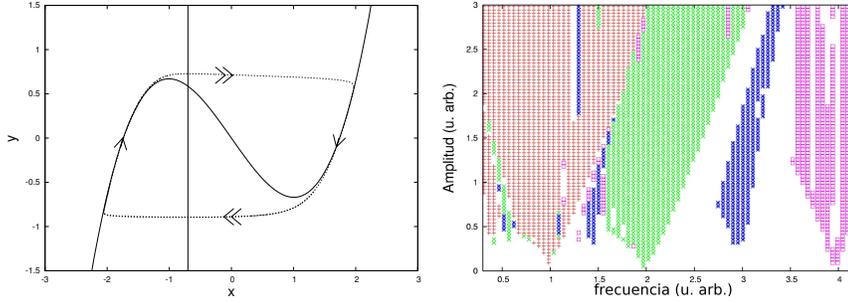


Figura 3.1: (Panel izquierdo): trayectoria en el espacio de fases para el oscilador de Van der Pol $\ddot{x} = -\mu(x^2 - 1)\dot{x} - x - a + A\cos(\omega_f t)$, $a = 0,7$. (Panel derecho): variando las amplitudes A y frecuencias de forzado ω_f , aparecen regiones de soluciones subarmónicas. En particular, graficamos las soluciones $1/1$ (rojo), $2/3$ (azul), $1/2$ (verde), $1/3$ (azul) y $1/4$ (rosa).

Hay diferencias, sin embargo, entre el diagrama de *lenguas de Arnold* para uno y otro caso (figuras 3.1 y 3.2) [HopIzh97, FGPV88]. Para los osciladores forzados, los extremos de las distintas lenguas de Arnold tocan el eje horizontal ($A = 0$) (fig. 3.2). En el diagrama correspondiente a los sistemas excitables (fig. 3.2), las lenguas están inclinadas de modo tal que nunca tocan el eje horizontal, como consecuencia de la *frecuencia natural infinita* del sistema excitable.

3.2. Invariantes topológicos

Los intentos por clasificar atractores extraños inspiraron invariantes de tipo métrico, dinámico y topológico. Los primeros incluyen una caracterización *dimensional* de varios tipos y funciones multifractales. Los exponentes de Lyapunov corresponden a la segunda familia. Todos ellos son invariantes ante transformaciones de coordenadas pero no ante el cambio en los parámetros de control, de manera que no son robustos ante perturbaciones de condiciones experimentales.

Las descripciones topológicas, en particular, permiten describir la organización de órbitas periódicas tridimensionales a través de números enteros o racionales. Para estos sistemas se pueden describir distintas clases de invariantes topológicos, entre ellos, los números de conexión (*linking numbers*) y las tasas de rotación relativa (*relative rotation rates*).

En esta sección se introducen las definiciones de estos invariantes y algunas de sus propiedades y algoritmos de construcción. En los siguientes capítulos presentaremos estos índices de dos maneras distintas. En primer lugar, para caracterizar fenomenológicamente propiedades espectrales de la voz en la búsqueda de una *huella vocal*. Y luego, en el capítulo de control motor neuronal, para comparar patrones temporales de presión con las soluciones de un modelo neuronal que permite generarlas sintéticamente.

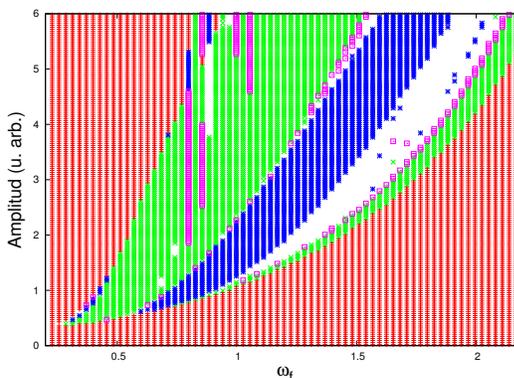


Figura 3.2: Estructura de soluciones subarmónicas estables para el sistema anterior, $\ddot{x} = -\mu(x^2 - 1)\dot{x} - x - a + A\cos(\omega_f t)$, en régimen excitable ($a = 1,35$). Las distintas lenguas corresponden a las soluciones subarmónicas con el mismo código de símbolos usado en la figura 3.1.

3.2.1. Tasas de rotación relativa

Estos índices fueron originalmente introducidos como invariantes topológicos para describir sistemas dinámicos bidimensionales periódicamente forzados [SolGil88], como los presentados en la sección 3.1. Sin embargo, pueden construirse *tasas de rotación relativa* para una gran familia de sistemas dinámicos tridimensionales autónomos en R^3 , o simplemente curvas cerradas para las que se pueda encontrar una *sección de Poincaré*. Informalmente, cuando la órbita tiene un “agujero” en medio, existe una familia de secciones de Poincaré que admiten el cálculo de tasas de rotación relativa.

Para construir estos índices, consideremos las curvas cerradas A y B , cuyos períodos son p_A y p_B , de forma tal que la curva A cruza la sección de Poincaré en p_A puntos a_1, a_2, \dots, a_{p_A} , y la B en p_B puntos b_1, b_2, \dots, b_{p_B} . Para cada par de condiciones iniciales (a_i, b_j) en la sección de Poincaré, el vector que conecta estos dos puntos vuelve a su configuración inicial después dar un número entero de vueltas en $p_A \times p_B$ períodos.

Las tasas de rotación relativa $R_{ij}(A, B)$ se definen como la rotación promedio, por período, de una curva sobre la otra, empezando por las condiciones iniciales a_i y b_j . Estos índices pueden ser expresados usando la integral

$$R_{ij}(A, B) = \frac{1}{2\pi p_A p_B} \oint \frac{n \times (\Delta \mathbf{r} \times d\Delta \mathbf{r})}{\Delta \mathbf{r} \Delta \mathbf{r}} \quad (3.3)$$

para cada una de las condiciones iniciales (a_i, b_j) , de modo que un total de $p_A \times p_B$ tasas de rotación relativa pueden construirse. Estas fracciones racionales no son típicamente todas iguales.

El cálculo de las tasas de rotación relativa para una órbita respecto de

sí misma, o *tasas de rotación propia* $R_{ij}(A, A)$, pueden calcularse imponiendo $R_{ii}(A, A) = 0$.

Hay algunas técnicas que permiten calcular estos números sin evaluar la integral de la ecuación 3.3. En el próximo capítulo mostraremos una, cuando construyamos tasas de rotación para funciones espectrales.

3.2.2. Números de conexión

Originalmente introducidos por Gauss para describir la organización de vórtices en el éter, el resultado de la integral

$$L(A, B) = \frac{1}{4\pi} \oint_A \oint_B \frac{(\mathbf{x}_A - \mathbf{x}_B) \times (d\mathbf{x}_A \times d\mathbf{x}_B)}{|\mathbf{x}_A - \mathbf{x}_B|^3} \quad (3.4)$$

calculada entre dos curvas cerradas de trayectorias \mathbf{x}_A y \mathbf{x}_B en R^3 sin puntos en común es un número entero invariante ante deformaciones de las curvas, siempre que no impliquen cruces entre ellas [Rol76].

No es necesario computar la integral definida en la ecuación 3.4; sobre una proyección bidimensional de las curvas, el siguiente algoritmo devuelve el valor de la integral:

- Se dibujan vectores tangentes a las dos curvas sobre el cruce.
- El vector superior se rota sobre el inferior en el sentido del menor ángulo entre ellos.
- Se asigna el valor $+1(-1)$ si la rotación es de “mano derecha” (“mano izquierda”).
- $L(A, B)$ tiene como valor la mitad de la suma de los valores de cruce a lo largo de las curvas.
- Los *números propios de conexión* $SL(A)$ se calculan como $2L(A, A)$.

En la figura 3.3 se ve un ejemplo del cálculo de los signos de los cruces para un nudo de trébol (*threefoil*).

Además, los índices $L(A, B)$ pueden calcularse directamente a partir de los $R_{ij}(A, B)$ (ec. 3.3) como

$$L(A, B) = \sum_{i,j} R_{ij}(A, B) \quad (3.5)$$

Para un análisis topológico es necesario contar con órbitas tridimensionales. Para el caso en que la dinámica sea originada por un sistema físico para el que se puedan medir tres series temporales distintas, en principio el problema está resuelto. Sin embargo, en muchos casos solamente una serie está disponible. El problema de mapear datos experimentales a trayectorias en \mathfrak{R}^n se llama *embebido*.

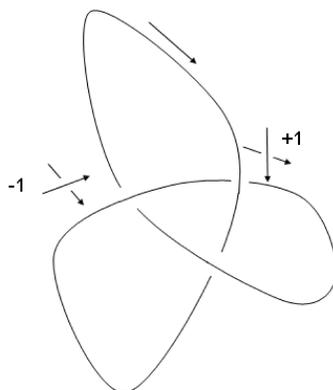


Figura 3.3: Ejemplos de cruces con signo para una órbita cerrada orientada tipo *threefoil*. Los números propios de conexión $SL(A)$ no siempre son invariantes topológicos, pero permiten caracterizar el nudo que forma la órbita.

3.2.3. Series embebidas

La elección del tipo de embebido particular depende de algún conocimiento del sistema que produce las series temporales. Aquí presentamos las técnicas de embebido más usuales.

Embebidos de retraso temporal

El tipo de embebido más común a partir de una serie $x(t)$ se construye según

$$\begin{aligned} x(t) \rightarrow \bar{y}(t) &= (y_1(t), y_2(t), \dots, y_n(t)) \\ y_k(t) &= x(t - \tau_k), k = 1, 2, \dots, n. \end{aligned} \quad (3.6)$$

Los parámetros τ_k son llamados *retrasos temporales*. Usualmente se expresan como múltiplos de un parámetro único τ , tal que $\tau_k = (k - 1)\tau$. Este tipo de embebido tiene la ventaja de poder construirse siempre, y de que todas las coordenadas tienen la misma relación señal a ruido.

Embebidos diferenciales

La manipulación de sistemas dinámicos y las ecuaciones de la física invita a pensar en términos de derivadas sucesivas. Esto sugiere un método para crear un embebido tridimensional a partir de datos escalares donde cada componente sea la derivada de la anterior,

$$\begin{aligned} x(t) \rightarrow \bar{y}(t) &= (y_1(t), y_2(t), y_3(t)) \\ \text{diferencial} &\rightarrow (x, dx/dt, d^2x/dt^2). \end{aligned} \quad (3.7)$$

Dentro de la familia de embebidos que se pueden aplicar en la construcción de órbitas en \mathfrak{R}^3 , éstos son los más relevantes para la exportación de la caracterización topológica de las funciones espectrales de la voz que estudiaremos a continuación.

En los próximos capítulos incorporaremos al problema de la voz algunas de las propiedades reseñadas aquí. El programa de caracterización topológica de los sistemas dinámicos que se desarrolló durante la década del 90 [SNM96, Gil98] armó a los dinamicistas con una batería de herramientas de clasificación y validación de modelos a partir de rígidas propiedades topológicas que se describen con números enteros o racionales. Esto inspirará la construcción de huellas topológicas para caracterizar los espectros de los sonidos voceados, redefiniendo el concepto de tradicional de ‘distancia’ entre espectros que se usa en la asignación de la identidad.

Capítulo 4

Huellas vocales y biometría

*‘Habré cambiado totalmente que el anciano
por la voz tan solo me reconoció’*

La casita de mis viejos
Cadícamo, Cobián, 1932.

En el capítulo 2 nos ocupamos de un sistema biométrico cuyos parámetros estaban relacionados con cantidades fisiológicas del aparato vocal. Los resultados obtenidos entonces sugieren que las características del tracto vocal son buenos indicadores biométricos de la identidad. Por otra parte, desde un punto de vista computacional, el costo para el ajuste de los parámetros es alto porque involucra la integración de ecuaciones diferenciales y el filtrado del flujo de aire. Estas observaciones invitan a una nueva interpretación del problema de la identificación de la identidad por la voz. Los contenidos de este capítulo fueron publicados en [TEM04].

Las aplicaciones tecnológicas asociadas a la voz pueden separarse en dos grandes grupos: el de *identificación del hablante* y el de *reconocimiento del habla*. Ambas aplicaciones comparten un aspecto central: tanto las propiedades que permiten discriminar una persona de otra como las que distinguen fonemas distintos se determinan en función de ‘distancias’ entre cualidades espectrales; en la sección 1.3.1 se derivaron algunas de las medidas más usadas en los sistemas de reconocimiento por la voz. Típicamente, estas medidas se inspiran en la percepción de los sonidos, que se inicia en el oído interno cuando las vibraciones de la cóclea producen una especie de transformada *mecánica* de Fourier, mapeando las frecuencias a lo largo de la partición coclear [Sm02]. En rigor, el problema perceptual es mucho más complejo: además de la decodificación de la identidad y del mensaje, tareas como la localización de la fuente sonora y la restauración de señales degradadas suceden simultáneamente durante la escucha de un discurso [Kee98, Rie98].

Desde el punto de vista tecnológico, las áreas de reconocimiento de la voz y de la identidad son opuestas. Para el reconocimiento del discurso, los siste-

mas se entrenan para independizarse del hablante, mientras que los esfuerzos para el reconocimiento de la identidad (en particular las aplicaciones *texto-independientes*) buscan independencia del contenido.

La estructura básica de los sistemas de reconocimiento del hablante consiste entonces en

- La definición de funciones o coeficientes espectrales apropiados, y
- La determinación de *umbrales de decisión* θ_0 a partir de los cuales aceptar o rechazar al usuario.

El problema se convierte entonces en el de la minimización de falsos positivos (aceptación de una identidad falsa) y falsos negativos (rechazo de una identidad verdadera) en función de θ_0 .

En cuanto al primer punto, entre las medidas de distorsión diseñadas para cuantificar similitudes entre dos funciones espectrales, las más difundidas calculan diferencias cuadráticas entre $H_1(f)$ y $H_2(f)$ y entre pendientes espectrales $\partial H_1(f)/\partial f$ y $\partial H_2(f)/\partial f$ (ecs. 1.11 y 1.12).

Una primera clave para la reinterpretación de las medidas espectrales queda sugerida por dos hechos notables. En primer lugar, el uso de funciones y sus derivadas sucesivas como medidas de distorsión evoca las técnicas de construcción de embebidos tridimensionales a partir de series temporales (sección 3.2.3). Y en segundo lugar, el uso de transformadas en el dominio de la frecuencia para caracterizar espectros (como los coeficientes *cepstrum*, sección 1.2 invita a pensar en los espectros como dichas series ‘temporales’. Sobre estas analogías proponemos en este capítulo un nuevo concepto de distancia entre espectros.

El propósito de este capítulo es doble; primero, la adaptación de las herramientas de la teoría de sistemas dinámicos a funciones espectrales de sonidos voceados. La descripción topológica resulta de importancia central para deshacerse de los umbrales de decisión. Por otra parte, reportamos un resultado empírico: hasta el grado de nuestra exploración, pueden definirse matrices unívocamente asociadas al hablante que las produjo. Esto nos permitirá definir el concepto de *huella vocal*.

4.1. Topología y voz

4.1.1. Reconstrucción espectral de sonidos voceados

En el primer capítulo se describieron algunas técnicas típicas de reconstrucción espectral a partir de series experimentales. En particular, la más ampliamente difundida es la *transformada rápida de Fourier*, o FFT.

Otras descripciones espectrales presentan ventajas respecto de la FFT. En particular, en la sección 1.2 se describió una técnica que permite reconstruir el espectro de potencias usando una serie de coeficientes d_0, d_1, \dots, d_m , $m \ll n$, a través de la función analítica continua y derivable definida por

$$H(f) = \frac{a_0}{1 - \sum_{k=1}^m d_k e^{ik2\pi f\Delta}}, \quad (4.1)$$

periódica en $[-1/2\Delta, 1/2\Delta]$. En la figura 4.1 se grafican los espectros de potencia en escala logarítmica $LH(f)$ definidas por

$$LH(f) \equiv \log|H(f)|^2 \quad (4.2)$$

correspondientes a varias enunciaciones de la misma vocal para dos personas distintas usando $m = 13$ coeficientes. Con este número de coeficientes, los espectros reconstruidos capturan las características principales de la envolvente espectral y de las pendientes de las formantes de los espectros.

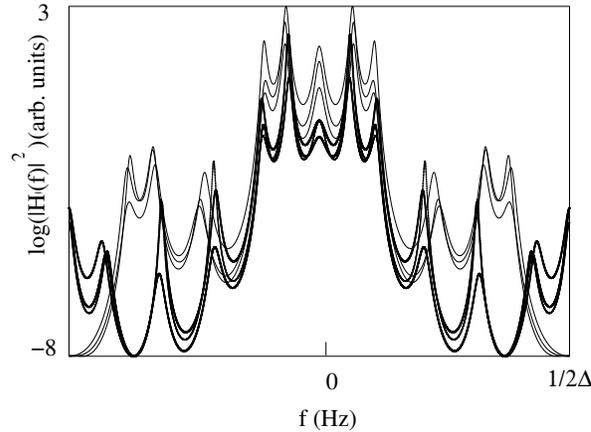


Figura 4.1: Tres ejemplos de $\log|H(f)|^2$ usando la aproximación de máxima entropía para dos hablantes distintos (en línea clara y oscura para cada uno). Más allá de la segunda formante, se distinguen fácilmente dos conjuntos de curvas. Los sonidos originales corresponden a pronunciaciones de la vocal [a] extraídas de muestras durante el habla normal. La tasa de muestreo es $r = 1/\Delta = 11,025\text{Hz}$.

Notablemente, las funciones resultan fácilmente clasificables en dos grupos con propiedades espectrales cercanas.

Para la tasa de muestreo elegida, $1/\Delta = 11,025\text{ Hz}$, el rango de frecuencia incluye típicamente las primeras 4 o 5 primeras formantes de los espectros. Esta elección no es arbitraria. En la figura 4.2 se muestra una serie de espectros calculados en el rango $0 \leq f \leq 22,050\text{ Hz}$ (panel izquierdo) y para el rango $0 \leq f \leq 5512\text{ Hz}$ (panel derecho). En la primera se nota una fuerte degradación de la repetitividad entre espectros para las frecuencias más altas. Esta degradación tiene dos orígenes. El primero está asociado al propio aparato vocal, que transmite menos información para las frecuencias altas, y el segundo es de orden tecnológico. En este trabajo usamos micrófonos comerciales, cuya respuesta en

frecuencia no se garantiza para las altas frecuencias. Sin embargo, en pruebas más cuidadosas, con micrófonos ATM 33a (tipo cardioide) en aislación acústica, también se observaron variaciones espectrales para alta frecuencia para distintas vocalizaciones del mismo sonido.

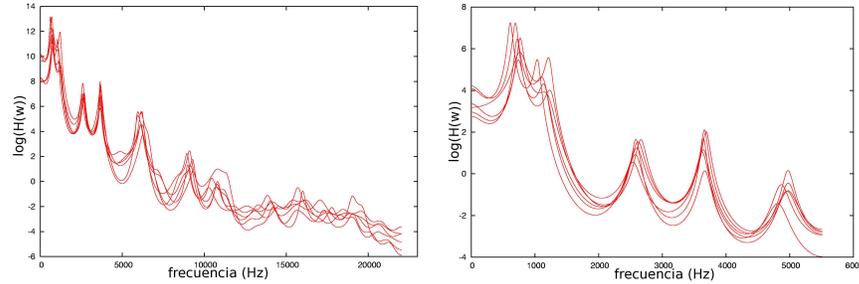


Figura 4.2: Espectros calculados con una tasa de muestreo de $r = 44100$ Hz (45 coeficientes lpc) y $r = 11025$ Hz ($m = 13$ coeficientes lpc) respectivamente. Para las resonancias correspondientes a frecuencias altas la repetitividad se degrada.

4.1.2. Construcción de matrices topológicas

Las funciones $LH(f)$ son simétricas respecto de $f = 0$. Por lo tanto, únicamente una mitad es relevante para nuestro análisis. Sumando una función lineal para anular la diferencia entre $LH(0)$ y $LH(1/2\Delta)$, y sustrayendo el promedio, construimos funciones $F(f)$ que resultan continuas y diferenciables al extenderlas al dominio real. En la figura 4.3 se muestran ejemplos de las funciones así construidas junto con una función, en línea punteada, que llamamos *función de referencia*.

Estas funciones pueden ser embebidas en el espacio de fases usando alguna de las técnicas descritas en 3.2.3. Por ejemplo, usando un retraso δ . Un ejemplo de la órbita resultante con $\delta = 40$ Hz puede verse en la figura 4.4.

Las órbitas así embebidas en el espacio de fases ($F(f), F(f - \delta), F(f - 2\delta)$) se enrollan alrededor de la línea $F(f) = F(f - \delta) = F(f - 2\delta)$. Queda por lo tanto bien definida una sección de Poincaré, que tomamos en el semiplano definido por $F(f) = F(f - 2\delta); F(f - \delta) \leq F(f - 2\delta)$.

En la sección 3.2 se describieron algunas de las herramientas topológicas usadas para caracterizar y clasificar algebraicamente las órbitas.

En este trabajo exploramos en particular el uso de *tasas de rotación relativa* y de *tasas de rotación propia*. Como la computación de las tasas de rotación propia se realiza respecto de la propia órbita, las diferencias de altura en las formantes se reflejan como índices distintos, que resultaron sensibles a las fluctuaciones naturales del espectro durante el habla.

Por otra parte, las tasas de rotación relativa muestran empíricamente una mayor cantidad de índices topológicos robustos, computados respecto de una

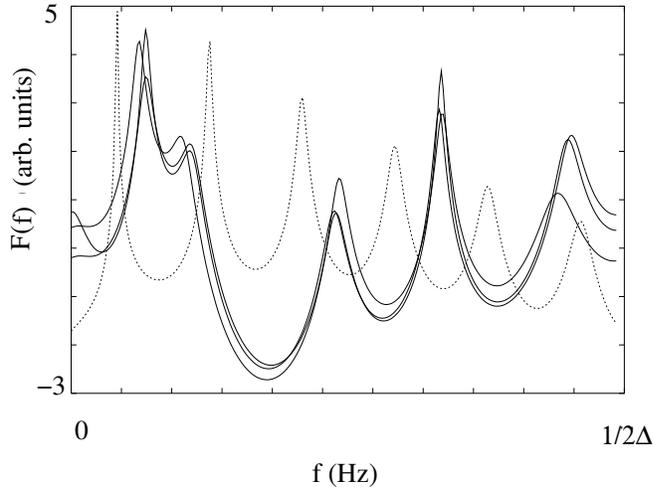


Figura 4.3: Funciones periódicas $F(f)$ construidas a partir de enunciaciones de un único hablante pronunciando repetidas veces la misma vocal (líneas sólidas) y una función de referencia (línea punteada). Estas funciones se construyen a partir de las funciones originales $LH(f)$ usando la mitad del período original.

órbita de referencia, graficada en la figura 4.3. El espectro sobre el que se construye esta órbita fue ya presentado en 2.1.2, donde se introdujo un *espectro de referencia* que corresponde a un tubo no articulado de cuarto de onda de longitud $L = 17,5$ cm, promedio del tracto vocal adulto. La elección de esta referencia está basada en la observación de que las geometrías de un tracto vocal real se pueden pensar como apartamientos de un tubo uniforme.

4.1.3. Índices y órbitas

En lo que sigue describiremos la caracterización de las órbitas usando tasas de rotación relativa. Llamamos *período* de una órbita al número de intersecciones que presenta con su sección de Poincaré, y a cada sección de la órbita entre intersecciones sucesivas *segmento*. Las rotaciones relativas de los espectros embebidos pueden ser calculados de la siguiente manera: asumiendo dos órbitas de períodos p_A y p_B , construimos una matriz de rotación $M \in \mathbb{Z}^{p_A \times p_B}$ cuyo elemento M_{ij} corresponde a la sumatoria de los cruces (con su signo) del i -ésimo segmento de la órbita A relativo al j -ésimo elemento de la órbita B . Los signos de los cruces pueden calcularse proyectando ambas órbitas en un espacio bidimensional. En dicha proyección, se dibujan los vectores tangentes sobre el cruce para cada segmento en la dirección del flujo. El vector superior se rota sobre el inferior, asignando un $+1$ o -1 para rotaciones de mano derecha o izquierda respectivamente.

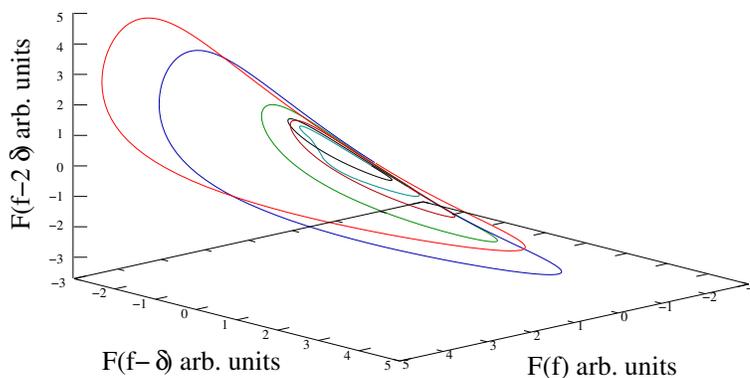


Figura 4.4: Embebido de retraso $\delta = 40$ Hz de la función $F(f)$ calculada a partir de una de las funciones espectrales de la figura 4.1. Los distintos tipos de línea corresponden a secciones de la órbita entre cruces de la sección de Poincaré.

La matriz resultante está relacionada con las tasas de rotación relativa a través de

$$R_{ij}(A, B) = \frac{1}{p_{APB}} \sum_{k=0}^{p_{APB}-1} M_{i+k, j+k} \quad (4.3)$$

tomando condiciones de contorno periódicas para la matriz. Las tasas de rotación relativa son el promedio de los elementos de matriz M_{ij} (ecuación 3.3) hasta repetir las condiciones iniciales (i, j) .

Por otra parte, cada uno de los elementos de la matriz M permite guardar la información de la rotación relativa de cada par de segmentos, y por lo tanto sirve para identificar variaciones espectrales para cada par de formantes de la referencia y la vocalización.

Para construir una huella vocal, comenzamos entonces por caracterizar un conjunto de K enunciaciones de una dada vocal para cada hablante. Para eso superponemos todas las matrices individuales correspondientes a cada sonido voceado y buscamos los índices coincidentes $M_{ij}^{(k)}$, $1 < k < K$ para las distintas enunciaciones. Llamamos *números robustos* a los índices que permanecen invariantes para el conjunto de K enunciaciones. El arreglo de números robustos en la matriz corresponde a una huella para la vocal elegida. En la figura 4.5 se ven algunos ejemplos de los arreglos resultantes para tres personas del mismo sexo y edades similares. Notablemente, el conjunto de números robustos es distinto en cada caso.

Llamamos *huella vocal* de un sujeto a la colección de las huellas de un dado conjunto de sus vocales.

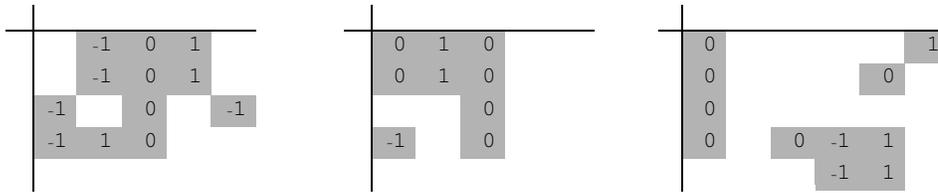


Figura 4.5: Huellas de voz para tres hablantes del mismo sexo y edades cercanas construidas a partir de segmentos cortos (~ 100)ms correspondientes a la vocal /a/ extraídos de 10 enunciaciones distintas, en distintas sesiones de enrolamiento.

4.1.4. Implementación del método topológico

Con estas definiciones implementamos un sistema de reconocimiento de la identidad de tipo *texto independiente* usando huellas topológicas.

La etapa de entrenamiento consistió en la creación del banco de usuarios registrados. Se construyó un banco de huellas vocales usando 10 segmentos (~ 100 ms) de distintas vocales tomados de textos leídos con registros de voz variados (distintas frecuencias fundamentales) y a distintas intensidades. A partir de los segmentos voceados se construyeron las matrices topológicas de cada vocal, se superpusieron y se construyeron las huellas usando los índices repetidos al menos el 90% de los casos.

El procedimiento de reconocimiento queda esquematizado en la figura 4.6. Una vez creado el banco de huellas vocales, se registraron nuevas muestras de voz tomadas de un texto y se construyeron las matrices topológicas de cada vocal (representadas como *candidatos* en la figura 4.6). Se aceptaron solamente las matrices que ajustaban perfectamente con las huellas vocales (en el esquema de la figura 4.6, la huella 1 para la matriz del candidato 1, y la huella 3 para el candidato 2.)

4.2. Resultados y discusión

En distintos conjuntos reducidos usados para las pruebas (~ 5 personas) la identificación fue unívoca, sin falsos positivos o negativos, para todos los casos.

Varios puntos fueron analizados para esta implementación. Por ejemplo, la elección de un subgrupo de los índices topológicos para las huellas vocales puede sugerir que en el proceso de reconocimiento se pierde información. Para verificar esta hipótesis, construimos un nuevo banco de datos, reemplazando la huella vocal de cada persona registrada por la colección completa de las matrices de rotación usadas para construirlas, de manera que toda la información topológica se mantuviera. Implementamos el método contando la cantidad de coincidencias entre los candidatos y las matrices de rotación del banco. El resultado fue

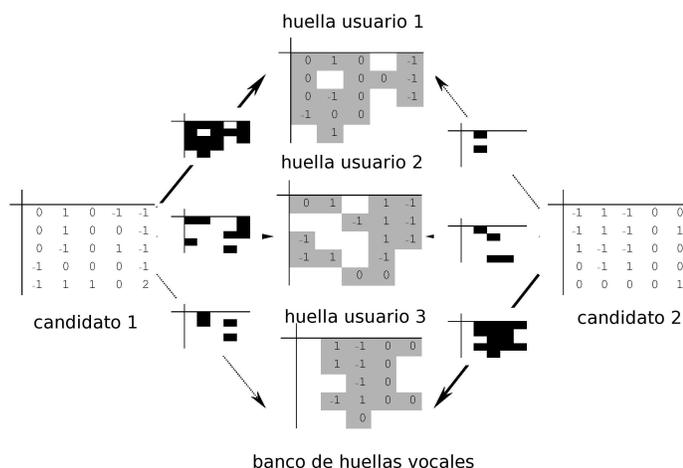


Figura 4.6: Esquema del proceso de reconocimiento: dos candidatos (de los cuales se muestran sólo una *huella voceada* de cada uno), se comparan con el banco de huellas. Un hablante es reconocido como un miembro del banco si el set de sus huellas coincide perfectamente con una huella registrada. Las áreas grises corresponden a las posiciones de las matrices que contienen números robustos. La identificación de un candidato como miembro del banco requiere que los números en esas posiciones para el candidato coincidan con los números robustos de la huella.

peor que en el caso de la huella vocal, ya que se encontraron varios positivos y negativos.

Concluimos que, hasta el grado de nuestra exploración, los números robustos parecen resaltar las características espectrales relevantes, mientras que los índices más variables, por su parte, no parecen guardar información alguna de la identidad, volviendo ruidoso al sistema.

Comparamos también la aproximación topológica con un método métrico simple, en el que la distancia cuadrática entre los espectros es usada como medida de distorsión (sección 1.3.1, ecuación 1.11). Los resultados se consideran aceptables debajo de un umbral optimizado para el banco de datos construido. En este caso, la huella vocal de cada persona registrada se reemplazó por las funciones espectrales $F(f)$ usadas para construir las matrices de rotación. Para este test también se verificaron falsos positivos y negativos, que no se encontraron con la implementación de la huella vocal.

La aproximación al problema biométrico de la voz presentado en este capítulo presenta varias ventajas respecto de los métodos usuales. En primer lugar, con una estrategia métrica (en el sentido de distancias entre espectros), debe definirse un umbral que típicamente depende del banco de datos usado. El uso

de huellas vocales construidas con números enteros introduce una estrategia novedosa, independiente del banco de voz utilizado y sin necesidad de umbrales de decisión.

Las implementaciones de la estrategia topológica, por otra parte, mostraron un costo computacional muy bajo. Una vez que la voz ha sido registrada, los segmentos voceados pueden ser extraídos fácilmente. Sus matrices de rotación relativa se construyen usando algoritmos de conteo de cruces muy simples [Gil98], y finalmente, las huellas vocales se calculan contando coincidencias en colecciones de matrices pequeñas. Por otra parte, una vez que el banco ha sido construido, toda la tarea de reconocimiento se limita a la comparación de matrices pequeñas como las que se ven en la figura 4.5.

En esta aproximación al problema, la construcción de un banco de vocales es un punto crítico. Exploramos el cambio en la cantidad de números robustos en la huella vocal en función del tamaño del conjunto de entrenamiento. Encontramos que para conjuntos de más de 10 vocales, el número de índices robustos converge aproximadamente a 8.

Comparamos también la huella vocal de un sujeto contra las matrices construidas con registros de voz durante un fuerte resfriado. Encontramos que la información en la matriz de números robustos se degrada sólo en las posiciones de matriz que corresponden a las formantes más altas, mientras que una gran parte de la huella vocal queda inalterada.

Para los sistemas dinámicos, los índices topológicos asociados a sus órbitas periódicas son invariantes. En nuestro trabajo, el hallazgo de índices topológicos muy estables puede sugerir la existencia de un sistema dinámico subyacente. Este punto de vista exige algunas precauciones. En primer lugar, se trata de un problema en el dominio de la frecuencia, y su análisis topológico representa un problema abierto. Los resultados de este trabajo son estrictamente fenomenológicos, e invitan primero a una fase de desarrollo sobre bases de datos masivas para establecer los alcances de su efectividad, comparando el método con técnicas más maduras.

El análisis desarrollado en este capítulo resulta cualitativamente distinto de los desarrollados con técnicas clásicas para atacar el problema del reconocimiento de la identidad por la voz, y permite sentar las bases para el desarrollo de una nueva tecnología [Pat03] capaz de generar soluciones biométricas independientes de cualquier umbral de decisión.

Capítulo 5

Control neuronal vocal

Después, Pan escuchó el viento pasar entre las cañas; creyó oír la voz triste de la ninfa y construyó un instrumento con cañas desiguales pegadas con cera, al cual decidió darle el nombre de siringe.

Diccionario de mitología griega y romana, Pierre Grimal.

En los primeros capítulos de esta tesis se discuten dos problemas del estudio de la voz humana. El modelado de la fisiología, por una parte, y las propiedades matemáticas de los espectros (topología de funciones espectrales) por otra. Ambos problemas se ordenan a un mismo objeto: sus posibles aplicaciones biométricas, es decir, la búsqueda de relaciones unívocas entre el registro de la voz y las particularidades del aparato fonador que lo genera. El capítulo final de esta tesis está dedicado, en cambio, al problema del control motor del sistema vocal.

En términos evolutivos, la generación de sonidos complejos no es una habilidad única del género humano. La comunicación a través del sonido es uno de los modos de interacción a distancia mejor establecidos y más variados en la naturaleza. Los insectos, por ejemplo, usan la vibración de parte o de todo el cuerpo para generar intensidad sonora suficiente para la comunicación. Los animales superiores presentan generalmente mayor especialización a través de un sistema dedicado a la generación de sonido. En particular, las aves cuentan con un aparato de fonación, la *siringe*, cuya estructura es análoga a la de la glotis en los humanos. En ella se aloja una estructura bilateral de tejido blando que es capaz de sostener oscilaciones mediante la transferencia de energía desde el flujo de aire, como en las cuerdas vocales.

Hay otro aspecto que distingue al habla de los sonidos producidos por otros organismos. Su control motor no es innato; las habilidades vocales se adquieren a través del aprendizaje. En la naturaleza, este fenómeno es extremadamente acotado. Hay evidencia de que algunos mamíferos superiores entrenan y corrigen sus patrones vocales [BraDou02]. Pero el caso de las aves importa especialmente aquí, porque así como existen especies que adquieren el canto, para otras la

actividad vocal es innata, y no exige más que un mínimo entrenamiento durante la etapa juvenil; luego, la percepción del propio canto es suficiente.

Las aves para las que el canto puede ser considerado una actividad aprendida requieren de un sistema de adquisición de canto [Not93], y constituyen una novedad evolutiva desarrollada independientemente en varios grupos de aves [BraDou02, Dub98].

En los humanos, el sistema neuromotor del habla es extremadamente complejo: además de la ejecución de los gestos de presión pulmonar, tensión de las cuerdas vocales y configuración del tracto vocal, el sustrato neuronal ejerce el control y la coordinación de estos gestos asistido por receptores de la acción mecánica y de la frecuencia fundamental por receptores auditivos [Tit93, MLG88]. Estos ingredientes evidencian un sofisticado programa para ejecutar y controlar el proceso del habla. El estudio del control motor en aves resulta especialmente pertinente aquí, porque además de presentar analogías básicas con la producción del habla, es más simple por su lejanía en la línea evolutiva, y representa una gran oportunidad para construir modelos apoyados en detallados estudios anatómicos y fisiológicos [Dub98, Not76].

La primera parte del capítulo está dedicado a una breve descripción de los puntos de encuentro entre el sistema neurovocal de las aves y del hombre y luego, sobre esta base, al desarrollo de modelos que permiten simular cantos a partir de *instrucciones neuronales* simples.

Finalmente, se sugieren posibles fuentes de complejidad vocal, con énfasis en el rol de la *interacción del sustrato neuronal con los órganos periféricos* para la generación de comportamientos acústicos complejos. Los resultados corresponden a las referencias [JTM05, ATM05, TMG06, TBSM05].

5.1. Actividad neuronal

En los procesos vocales, como en cualquier otra actividad motriz, los músculos están controlados por patrones de actividad nerviosa que emerge de la interacción entre miles de células específicas (neuronas) que funcionan como unidades excitables, organizadas en conjuntos bien definidos llamados *núcleos*. Los núcleos a su vez están interconectados por *interneuronas* y organizados en una jerarquía bien definida; desde los núcleos más bajos, encargados de enervar a los músculos, hasta los implicados en tareas de procesamiento de la información sensitiva, de aprendizaje, de control.

Esta organización sugiere que un nivel de descripción posible para el problema es el de la *actividad media* del conjunto de neuronas que compone cada núcleo, en lugar de la actividad individual de las neuronas.

En las células, el medio intracelular está separado del exterior por una membrana que permite el intercambio de moléculas. El control del volumen celular implica el transporte activo de iones entre el interior y el exterior, lo que resulta en una diferencia de potencial a través de la membrana. La regulación del

potencial de membrana mediante el control de los canales iónicos es una de las funciones celulares más importantes. Muchas células, como las musculares y las neuronas, usan el potencial de membrana para la transmisión de señales. Toda la operación del sistema nervioso y la contracción muscular, por ejemplo, dependen de la generación y propagación de estas señales eléctricas, que consisten en pulsos (*spikes*) acotados en el espacio y en el tiempo que tienden a *depolarizar* o *hiperpolarizar* la membrana, según reduzcan o aumenten la diferencia de potencial intracelular respecto del exterior.

Según su acción sobre las que se conectan con ellas, las neuronas pueden clasificarse en dos categorías: las excitatorias, que favorecen la depolarización de la célula que recibe la señal (postsináptica), y las inhibitorias, que la hiperpolarizan. Una interacción excitatoria facilita la aparición de un pulso en la célula postsináptica, mientras que una inhibitoria lo reprime.

Existe una gran variedad de modelos para la actividad neuronal. En términos de la tasa de producción de pulsos, las neuronas responden a los estímulos recibidos (externos y de otras neuronas) por encima de un umbral a través de una función saturable (S).

Supongamos un conjunto N_e de neuronas excitatorias y N_i de neuronas inhibitorias interconectadas en un núcleo. La evolución temporal de la tasa de actividad para las neuronas excitatorias (e) o inhibitorias (i) puede escribirse como [SchWag90]:

$$\frac{de_k}{dt} = -e_k + S \left[a_e \left(\frac{1}{N_e} \sum_l^{N_e} u_{kl} e_l - \frac{1}{N_i} \sum_l^{N_i} v_{kl} i_l - \theta_k^e + p_k \right) \right] \quad (5.1)$$

$$\frac{di_k}{dt} = -i_k + S \left[a_i \left(\frac{1}{N_e} \sum_l^{N_e} w_{kl} e_l - \frac{1}{N_i} \sum_l^{N_i} z_{kl} i_l - \theta_k^i \right) \right], \quad (5.2)$$

con $k = 1, 2, \dots, N_e, N_i$ en las ecs.5.1 y 5.2 respectivamente.

Cada célula responde a la suma de las actividades que recibe a través de una función no lineal saturable. A lo largo de este trabajo usaremos la función $S(x) = 1/(1 + e^{-x})$ [HopIzh97].

Las constantes $\theta_k^{e,i}$ representan umbrales para cada unidad, y p_k los estímulos externos. Los parámetros u_{kl} , v_{kl} , w_{kl} , y z_{kl} describen las conexiones entre unidades: por ejemplo, u_{kl} describe la conexión excitatoria entre la k -ésima ($k = 1, \dots, N_e$) y la l -ésima unidad.

Definiendo los coeficientes promedio $c_{ij} = (N_i N_j)^{-1} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} x_{kl}$ para los distintos parámetros de conectividad $x = u, v, w$ y z , las ecuaciones para la dinámica de los promedios $E(t) = N_e^{-1} \sum^{N_e} e_i(t)$ e $I(t) = N_i^{-1} \sum^{N_i} i_i(t)$ definidas por

$$\frac{dE}{dt} = -E + S [a_e (c_{ee} E + c_{ei} I + P^e - \Theta^e)] \quad (5.3)$$

$$\frac{dI}{dt} = -I + S [a_i (c_{ei}E + c_{ii}I - \Theta^i)]. \quad (5.4)$$

resultan válidas para describir la actividad promedio de cada población bajo la condición de que las conectividades individuales sean pequeños apartamientos de las conectividades promedio, $x_{kl} = c_{ij} + \delta x_{kl}$.

Se investigó la validez de esta aproximación comparando simulaciones numéricas para el promedio de las actividades (ecs. 5.1 y 5.2) y las actividades promedio (ecs. 5.3 y 5.4) sobre poblaciones de 100 unidades excitatorias y 100 inhibitorias con 25000 conexiones, estimulando los sistemas con funciones periódicas.

Las conectividades se calcularon para variables aleatorias de distribución uniforme $\delta x_{kl} \sim U[-\alpha c_{ij}, \alpha c_{ij}]$, $\alpha \in (0, 1)$.

El acuerdo entre ambas dinámicas se mantiene cercano hasta valores tan grandes como $\alpha = 0.6$. También se exploró el efecto de inhomogeneidades espaciales en las unidades estimuladas. En estas condiciones, aún con el 10% de unidades estimuladas en el sistema extenso, los campos promedio son bien aproximados por el sistema promediado.

La similitud entre las dinámicas de ambos modelos y el verdadero sistema motor no puede mantenerse más allá de características generales. En este trabajo se busca una mínima receta computacional para explorar las respuestas de ‘poblaciones’, es decir, de actividad global bajo la acción de ritmos simples, de manera que un ritmo básico pueda generar distintos patrones de actividad neuronal.

En lo que sigue describiremos la actividad de los núcleos motores con las ecuaciones promedio 5.3 y 5.4.

Tanto en aves como en mamíferos, los núcleos implicados en el control vocal presentan una organización jerárquica. Según el esquema de la figura 5.1, el camino de control motor de las aves va desde el núcleo HVC (*High vocal center*) al RA (*Robustus archistriatalis*), pertenecientes al telencéfalo. En RA hay neuronas que proyectan a los núcleos medulares nXIIts (que controla el aparato vocal) y otras que proyectan a los núcleos respiratorios RAm (*retroambigualis*) y PAm (*parambigualis*). RA cumple aquí el rol de caja articuladora, donde las instrucciones que llegan de HVC se separan hacia los núcleos siríngeos y respiratorios para producir las vocalizaciones.

Debemos distinguir aquí dos grupos de aves: los *oscinos*, que cuentan con núcleos telencefálicos descritos antes, que controlan el programa del canto y del aprendizaje, y los *no-oscinos* que sólo cuentan con los núcleos motores y premotores (vocales y respiratorios) que interactúan con los órganos ejecutores de los gestos del canto.

Esta separación natural en dos grandes grupos de aves cuya estructura neuronal es muy distinta, y son capaces de generar patrones acústicos complejos resulta un punto de partida óptimo para ensayar una pregunta a la que se dedica, implícitamente, el contenido de este capítulo: ¿Cuál es punto de surgimiento de la complejidad de los patrones acústicos?

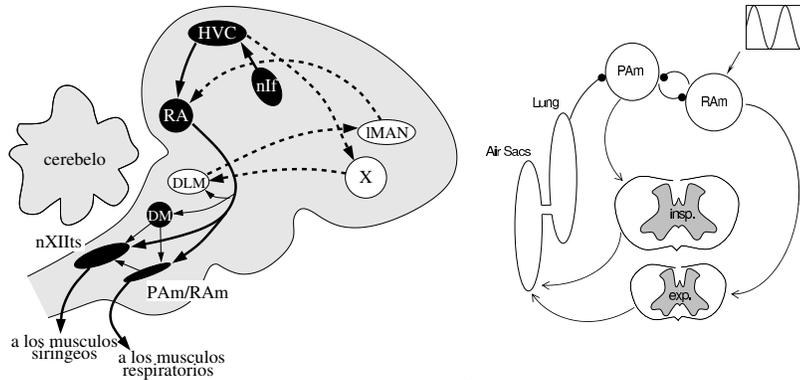


Figura 5.1: Panel izquierdo. Esquema de los principales núcleos implicados en el control motor del canto en oscinos. En el telencéfalo, las instrucciones bajan desde los núcleos HVC a RA, donde se separan entre las que controlan los núcleos medulares vocales (nXIIIts, que enerva músculos de la siringe) y respiratorios (RAm y PAm, que controlan la inspiración y la expiración). Panel derecho. Los *suboscinos* no cuentan con los núcleos telencefálicos, aunque comparten con aquel la estructura medular cuyas conexiones inhibitorias (círculos) y excitatorias (flechas) se indican en el esquema.

5.2. Neuromecánica vocal

Durante el habla, el sistema neuronal debe controlar esencialmente tres gestos anatómicos: la presión pulmonar, la tensión de las cuerdas vocales y la configuración de los articuladores del tracto vocal. En el canto de las aves, el último gesto es menos importante. En esta primera sección comparamos la anatomía vocal y respiratoria de humanos y de aves, para después introducir su control neuronal.

5.2.1. Siringe y glotis

Durante el habla, aún sin un contorno melódico prescrito, el hablante debe conformar un patrón de entonación para establecer una comunicación efectiva. Esta sintonía está determinada en parte por la lingüística. Por ejemplo, el incremento en la frecuencia fundamental al final de una interrogación o la caída al final de una frase, forman parte del código lingüístico [Tit93, Fla72, Rab93].

Los músculos que enervan la laringe (sección 1.1.1) estiran o relajan las fibras de las cuerdas vocales, modificando su tensión efectiva y la correspondiente frecuencia natural de las oscilaciones. La longitud de las cuerdas vocales constituye la principal variable geométrica para el ajuste de la frecuencia fundamental.

El control muscular de la longitud de las cuerdas a través del sistema de los músculos laríngeos es muy intrincado (figura 1.1) y puede conducir a una gran variedad de configuraciones musculares redundantes en términos de la tensión

efectiva resultante [Tit93].

En las aves, en cambio, la frecuencia fundamental se controla principalmente a través del músculo *siringealis ventralis* (vS) [GolSut96, GolSut96b]. Las diferencias anatómicas entre distintas taxonomías son enormes: hay aves que controlan una sola siringe (típicamente traqueal), mientras que otras controlan dos siringes bronquiales [Ele03]. En la figura 5.2 se esquematizan las siringes.

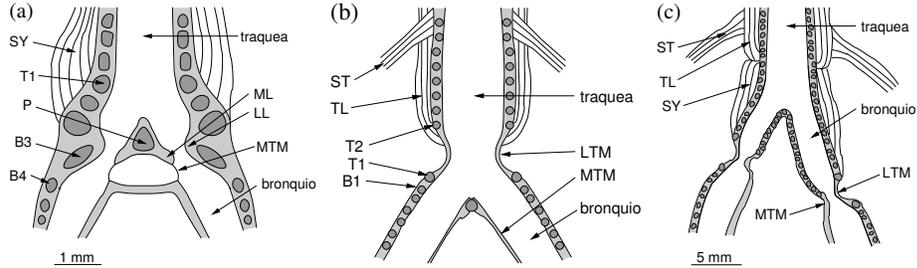


Figura 5.2: Tres tipos de órgano vocal en aves: siringe traqueobronquial (a), traqueal (b) y bronquial (c). El mecanismo de producción de sonido es, en todos los casos, análogo al del habla humana, generado por las oscilaciones autosostenidas de las membranas laterales, cuya tensión es controlada por un complejo sistema de músculos descritos en la figura 1.1.

A pesar de estas grandes diferencias anatómicas, todas las siringes contienen una estructura análoga a las cuerdas vocales en humanos, los *labia*. El modo en que los *labia* producen sonido en las aves es esencialmente el mismo con que las cuerdas vocales lo hacen en el caso del habla [GolSut96], mediante la interrupción cíclica del paso del aire a través de las membranas.

A partir de esta observación, el modelo de oscilación de cuerdas vocales estudiado en la sección 2.1.1 dio lugar a un modelo dinámico para el órgano vocal de las aves [Gar01].

La dinámica del modelo para las cuerdas vocales presentado en las ecuaciones (2.8) está caracterizada por el encendido de las oscilaciones a través de una bifurcación de Hopf. Para modelar la siringe se ha propuesto un modelo bidimensional que captura la transferencia de energía desde el flujo de aire al tejido mediante la competencia entre la presión del aire $p(t)$ y la disipación de las membranas, caracterizada por el coeficiente b . La tensión variable de los *labia* está caracterizada por la función $k(t)$. La ecuación dinámica de la siringe resulta

$$m\ddot{z} - [p(t) - b]\dot{z} + k(t) + cz^2\dot{z} = 0, \quad (5.5)$$

con m la masa efectiva de los *labia*, cuya separación de la posición de equilibrio queda descrita por la variable z , y c un coeficiente de disipación no lineal que da cuenta del choque de los *labia*. Esta es una ecuación que presenta sintéticamente los procesos presentados en el capítulo 2, es decir, la competencia entre

la disipación del tejido caracterizada por el coeficiente b y la energía entregada por el flujo de aire $p(t)$. El coeficiente de restitución $k(t)$ es en este caso explícitamente dependiente del tiempo.

En este esquema, los gestos necesarios para la vocalización quedan definidos con la tensión de los labia $k(t)$ y la presión de los sacos aéreos $p(t)$. Este modelo ha sido verificado mediante la comparación del canto real con canto sintetizado con el modelo alimentado con datos electromiográficos para $p(t)$ y $k(t)$ registrados simultáneamente con el canto [MinGar03].

Los músculos siríngeos están controlados por el núcleo nXIIIts (*hypoglossal tracheosiringeal*). Modelamos su actividad promedio n usando

$$\tau \dot{n} = -n + S(-E - I_2 + \text{instrucciones}). \quad (5.6)$$

Las *instrucciones* representan la actividad de núcleos superiores (figura 5.1) que enervan el nXIIIts, que también recibe actividad de centros respiratorios (I_2) [SWM03]. La variable E representa un nivel de actividad basal del núcleo.

Asumimos una relación lineal $k(t) \propto n(t)$, de manera que la acción de los músculos de la siringe resulte proporcional a la actividad neuronal del núcleo que los enerva.

5.2.2. Pulmones y sacos aéreos

En los mamíferos, los pulmones son los órganos responsables del intercambio de gases en la sangre a través las vías aéreas, un sistema de tubos ramificados cada vez más estrechos y cortos, a través de los cuales el movimiento de los gases se produce por simple difusión. El sistema respiratorio y el vocal están de este modo interconectados: el flujo de aire que se genera en el proceso respiratorio es aprovechado en las vías aéreas superiores (laringe), para generar el movimiento de las cuerdas vocales. En las aves, en cambio, hay otra estructura responsable de la ventilación de los pulmones, los *sacos aéreos*.

Comparando los caminos neuronales respiratorios de aves con los de mamíferos, parece existir un patrón similar de grupos de células y proyecciones que sugieren múltiples equivalencias entre ambas estructuras neuronales, con núcleos funcionalmente equivalentes [Dub98]. Sin embargo, algunas diferencias merecen apuntarse. Mientras que en mamíferos, según la descripción seguida en 1.1.1, los mecanismos expiratorios dependen esencialmente de la restitución mecánica del volumen pulmonar por la caja torácica, en las aves este control es activamente gobernado por el núcleo RAm (*nucleus retroambigualis*) [SWM03]. Otro núcleo específico, el PAm (*nucleus parambigualis*), enerva los músculos reponsables de la inspiración.

En este trabajo proponemos una dinámica para los sacos aéreos análoga a la utilizada para el diafragma en mamíferos (ecuación 1.1). Los núcleos que controlan la actividad inspiratoria I_1 (PAm) y expiratoria I_2 (RAm) presentan una interacción mutuamente inhibitoria [SWM03]. El modelo resulta

$$m\ddot{x} = -\kappa x - \mu\dot{x} + \alpha I_1 - \beta I_2, \quad (5.7)$$

$$\tau_1 \dot{I}_1 = -I_1 + S(-E_1 - aI_2 + bI_1 - f(x)), \quad (5.8)$$

$$\tau_2 \dot{I}_2 = -I_2 + S(-E_2 - cI_1 + dI_1 + \text{instrucciones}). \quad (5.9)$$

Las actividades basales se representan mediante las variables E_1 y E_2 para el PAm y RAm respectivamente.

En este caso, la variable x representa el desplazamiento de los sacos aéreos respecto de su posición de equilibrio a presión atmosférica. Para relacionar la dinámica de los sacos aéreos con la presión usamos la misma aproximación desarrollada en la sección 1.1.1, $p \propto -x$.

Existe también un mecanismo de realimentación desde la periferia. En particular, la función $f(x) = 9x^3/(1+x^3)$ representa la reacción de los sensores de CO_2 inhibiendo en la actividad responsable de la inspiración.

Un esquema de los elementos y las conexiones del modelo se muestra en la figura 5.1.

Las diferencias principales entre el sistema vocal humano y el de las aves se manifiestan comparando el sonograma de la figura 1.3 durante la pronunciación de una palabra y los de la figura 5.3 que corresponden a cantos de aves filogenéticamente lejanas.

Durante el habla, los sonidos voceados están caracterizados por bandas de frecuencias aproximadamente estacionarias. La estructura armónica contiene gran parte de la información, controlada por la configuración del tracto vocal. Por otra parte, el contorno melódico también codifica información lingüística, y depende del control de la tensión de las membranas vibrantes, como en las aves. En las aves, el canto está caracterizado por una estructura rítmica de distintas sílabas caracterizadas por una traza continua de frecuencia fundamental variable.

5.3. Subarmónicos y complejidad vocal

En la sección anterior construimos un modelo completo para la actividad vocal de las aves que depende exclusivamente de instrucciones de los núcleos neuronales: la ecuación 5.6 se usa para generar los gestos de tensión $k(t)$, y las ecuaciones 5.7, 5.8 y 5.9 definen los gestos respiratorios a través de la presión $p(t)$. Estos son los dos gestos que controlan la siringe (ecuación 5.5), de modo que el modelo completo permite generar vocalizaciones sintéticas a partir de parámetros neuronales.

El modelo respiratorio y vocal presentado constituye un modelo mínimo de control motor para el canto compartida por todos los grupos de aves. Los oscinos, sin embargo, tienen núcleos neuronales superiores de control motor. En el paradigma actual, la acción de estos centros está relacionada con la complejidad

del canto. En esta sección mostraremos que la estructura básica descrita alcanza para explicar la compleja estructura rítmica en aves de los dos subgrupos, reforzando la idea de que la complejidad puede asociarse a sustratos neuronales simples gobernados por ritmos básicos.

En la figura 5.3 se muestran los sonogramas de dos aves filogenéticamente lejanas; el panel superior corresponde al sonograma del canto a dúo de una pareja de horneros (*Furnarius Rufus*), y el inferior al de un canto típico de canario (*Serinus Canaria*).

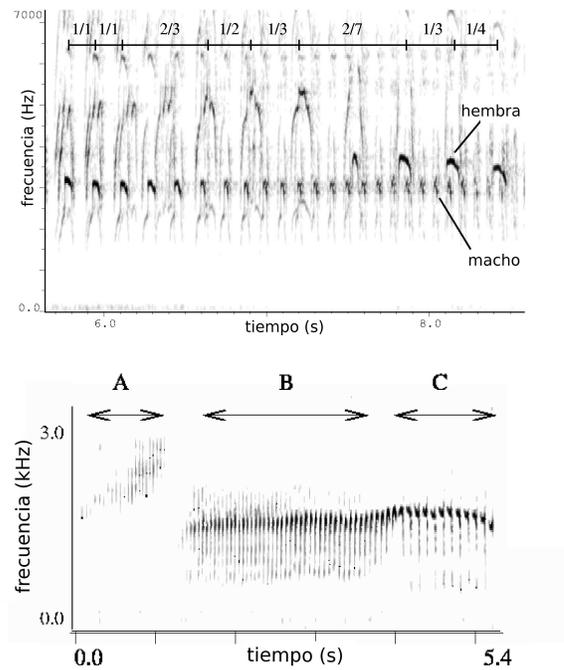


Figura 5.3: Sonogramas de dos cantos de aves distintas. Arriba, un dúo de horneros, que presenta una compleja trama de coincidencias entre sílabas del macho y de la hembra. Abajo, el sonograma del canto típico de un canario.

Suboscinos

La figura 5.3 muestra un sonograma típico de un dúo de horneros (*Furnarius Rufus*). Las trazas continuas del sonograma representan las variaciones de la frecuencia fundamental en cada sílaba. Las sílabas de frecuencia más baja, que conforman un canto simple de frecuencia intersilábica creciente, corresponden al macho. La hembra, por su parte, muestra un canto mucho más complejo, que caracterizamos aquí por su estructura de coincidencias con sílabas del macho. En la figura 5.3 las fracciones $r_{ap} = p/q$ se definen con el número p de sílabas de la hembra y el número q de sílabas del macho entre coincidencias consecutivas.

Una base de datos construida para 20 duetos se sintetiza en la figura 5.4. Cada duo corresponde a una altura distinta en el eje vertical. Cada fracción de la serie r_1, r_2, \dots, r_n de cada duo r_{ap} se localiza en el período intersilábico del macho donde ocurre (eje x), y su valor está asociado a un símbolo distinto. Los duetos fueron organizados para obtener una interfaz suave entre las dos regiones mayores, $r_{ap} = 1/3$ y $r_{ap} = 1/4$. Notablemente, este procedimiento deja regiones bien definidas y permite organizar los dúos en términos de la estructura rítmica, independiente de las características acústicas de cada sílaba particular.

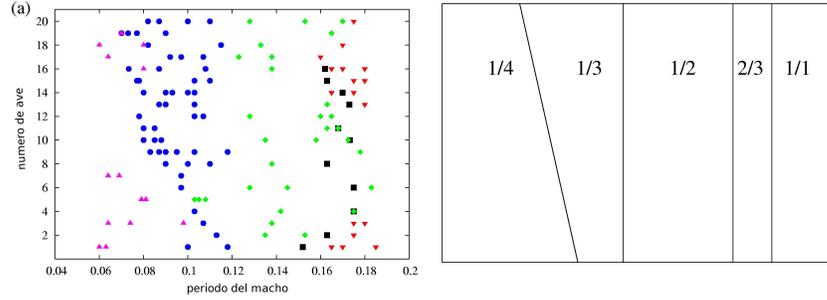


Figura 5.4: Organización rítmica del canto a duo de los horneros. (a) Un esquema de clasificación para las coincidencias r_{ap} intersilábicas de cada uno de los duos, ordenados en el eje vertical. Cada sílaba de la hembra se representa por un símbolo distinto, ubicado en el período correspondiente del macho. El símbolo ∇ para $r_{ap} = 1/1$, \diamond para $r_{ap} = 1/2$, \circ corresponde a $r_{ap} = 1/3$, \triangle para $r_{ap} = 1/4$ y para $r_{ap} = 2/3$. Con este esquema quedan definidas distintas regiones para cada enganche. A la derecha, orden esquemático de las regiones de enganche.

Para modelar estos datos usamos el sistema respiratorio desarrollado antes (ecuaciones 5.7, 5.9 y 5.8). Asumimos que RAm es enervado por el núcleo *dorsomedial nucleus intercolliculum* (DM), que se supone el más alto en el camino motor de los suboscinos. A pesar de que los mecanismos que relacionan el camino auditivo y el motor no son conocidos, el núcleo DM es adyacente al sistema auditivo [Wil97b], y asumiremos que tiene una actividad sensible a la audición.

$$m\ddot{x} + kx + \mu\dot{x} = 3I_1 - 4I_2 \quad (5.10)$$

$$\dot{I}_1 = 30[-I_1 + S(-1,43 - 18I_2 + 2I_1 - f(x))] \quad (5.11)$$

$$\dot{I}_2 = 30[-I_2 + S(-1,43 - 18I_1 + 2I_2 + A\cos(\omega t))] \quad (5.12)$$

Matemáticamente, representamos las instrucciones del DM sobre RAm (ecuación 5.12) usando una función armónica inducida por la actividad auditiva. Esta actividad representa el promedio de actividad de toda la población neuronal DM e identificamos la frecuencia ω como la frecuencia intersilábica del macho.

Sin forzado del DM, este sistema presenta excitabilidad (sección 3.1). Esta condición depende esencialmente de la interacción de los sacos aéreos y el sus-

trato neuronal. La respuesta subarmónica de este sistema presenta la estructura de lenguas presentada en la figura 5.5.

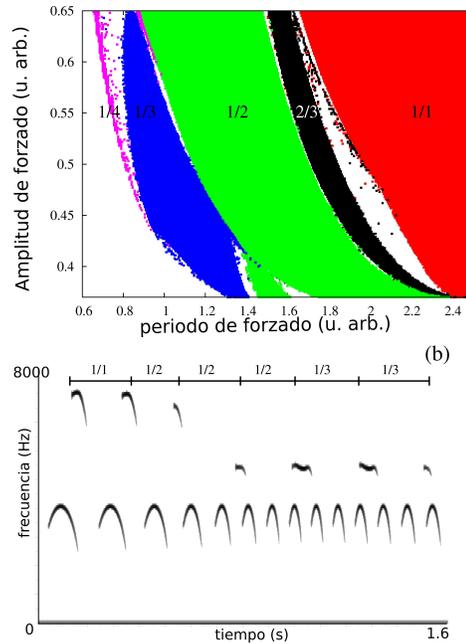


Figura 5.5: Arriba: el modelo descrito por las ecuaciones 5.10, 5.11 y 5.12 presenta la estructura de lenguas en términos de las variables $(2\pi/\omega, A)$ que caracterizan al forzado. Abajo: sonograma para la simulación del dúo alimentando el modelo de la siringe (ec. 5.5) con la salida del modelo respiratorio.

Notablemente, el orden relativo de estas lenguas coincide con el orden obtenido para las fracciones r_{ap} (figura 5.4), que de esta manera pueden identificarse con los números de rotación de las órbitas del sistema respiratorio. Es decir, los patrones respiratorios se organizan como las soluciones subarmónicas de un sistema no lineal forzado.

La identificación de las fracciones r_{ap} y los números de rotación r del sistema dinámico debe hacerse con cuidado. Para la construcción de la figura teórica se usaron soluciones después del transitorio, mientras que para el caso real, cuando la frecuencia del macho está incrementándose continuamente, los transitorios no pueden descartarse. El error cometido al identificar las fracciones r_{ap} de coincidencia con los números de rotación es inversamente proporcional a la duración del transitorio, por lo que el acuerdo encontrado sugiere que el sistema respiratorio de la hembra debe presentar un comportamiento altamente disipativo.

Finalmente, hemos sintetizado un dúo alimentando el modelo de la siringe con los patrones respiratorios generados por el modelo. El sonograma resultan-

te se puede ver en la figura 5.5, que reproduce las características rítmicas y acústicas observadas en cantos reales.

Los resultados de esta sección describen un posible mecanismo capaz de crear patrones complejos a partir de un sustrato neuronal simple en aves que no presentan núcleos motores telencefálicos. En este marco, la complejidad rítmica proviene de la interacción entre los centros neuronales respiratorios y los órganos respiratorios, sin intervención de una compleja estructura neuronal, y su estructura rítmica es compatible con la de un sistema excitable de baja dimensión. Apoyados en estos resultados, estudiamos la estructura rítmica de los canarios, que pertenecen al grupo de los oscinos, y que comparten la estructura respiratoria básica utilizada aquí para explicar la estructura rítmica de los duos.

Oscinos

En el inicio de este capítulo se discutieron diferencias esenciales entre los sistemas de control motor en oscinos y suboscinos. Aquel grupo se distingue por la rica estructura de núcleos telencefálicos, que se creen ligados a la emergencia de patrones complejos de canto [McCKon81]. Los canarios (*Serinus canaria*) pertenecen a este grupo de aves. El canto está formado por una larga secuencia de sílabas distintas, cada una de las cuales se repite un número variable de veces. El sonograma de la figura 5.3(b) presenta tres regímenes distintos (A, B y C) caracterizados por diferentes períodos silábicos. La serie temporal de la presión de los sacos aéreos medidos simultáneamente con el canto correspondiente a este sonograma (métodos B.1) se muestra en la figura 5.6(a).

Apoyados en los resultados obtenidos para suboscinos, proponemos que el sustrato respiratorio es suficiente para generar los patrones observados en oscinos. En este caso usamos

$$m\ddot{x} + kx + \mu\dot{x} = 2I_1 - I_2, \quad (5.13)$$

$$\tau\dot{I}_1 = -I_1 + S(E_1 - I_2 - f(x)), \quad (5.14)$$

$$\tau\dot{I}_2 = -I_2 + S(E_2 - I_1 + A \cos(\omega t)), \quad (5.15)$$

con parámetros ajustados para generar los patrones de presión sintéticos (figura 5.6b) para tres frecuencias ω distintas y amplitud constante A . El término de forzado armónico, en este caso, debe entenderse como una aproximación a los distintos ritmos silábicos impuestos desde el núcleo telencefálico RA.

Más allá de las similitudes entre las series experimentales y sintéticas, se embobieron las series temporales y se computaron los números de conexión para cada régimen (ver secciones 3.2.3 y 3.2). El propósito de este análisis es el de proveer una forma cuantitativa de contrastar los datos experimentales con el modelo. Los segmentos extraídos de los regímenes A y B no se conectan entre sí, mientras que los del régimen C se conectan con los segmentos del régimen B. la

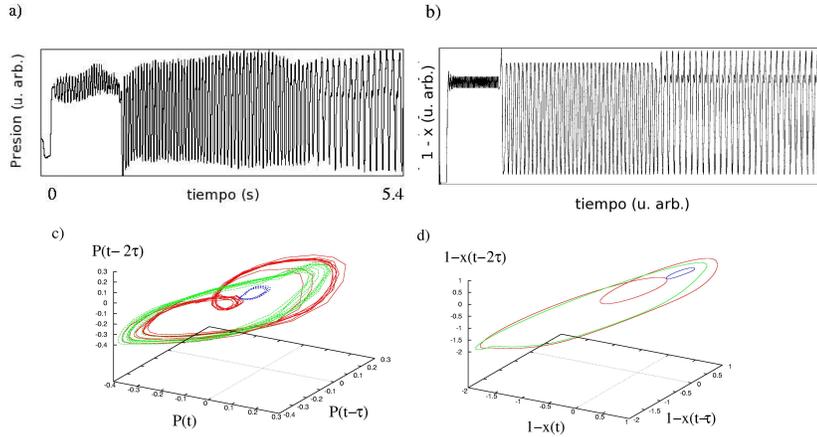


Figura 5.6: (a) Serie experimental de la presión de los sacos aéreos correspondiente al sonograma de la figura 5.3 para *Serinus Canaria*. (b) Serie numérica correspondiente al modelo de ecs. 5.13, 5.14 y 5.15 integrados con $m = 0,5$, $\tau = 1,0$, $k = 1,0$, $\mu = 5,0$, $E_1 = -1,3$ y $E_2 = -1,5$. (c) Segmentos embebidos con un retraso temporal de 0.3 s. La línea punteada (azul) corresponde al patrón A, la rayada (verde) al tipo B y la línea continua (rojo), al C. (d) Segmentos embebidos para la serie sintética, con el mismo código de color que en (c).

organización topológica en ambos casos es idéntica: $L(A, B) = 0$, $L(A, C) = 0$ y $L(B, C) = 1$.

El análisis se extendió a un conjunto de más de 100 tipos de patrones de presión para un conjunto de seis pájaros distintos. Los patrones se clasificaron de acuerdo a su contenido espectral. Para ilustrar el método utilizado, analizamos los regímenes de la figura 5.6(b).

Tanto el tipo A como el B presentan un sólo pico a 25 Hz y 20 Hz respectivamente. El C, en cambio, presenta dos picos: uno a 23 Hz y otro a 11.5 Hz. Estas transiciones admiten dos interpretaciones. O bien el ave cambia dramáticamente la frecuencia de oscilación de B ($f_0 \sim 20$ Hz) a C ($f_0 \sim 11,5$ Hz), o bien la cambia suavemente desde $f_0 \sim 20$ Hz a $f_0 \sim 23$ Hz, entre las que el sistema respiratorio sufre una duplicación de período, compatible con el modelo propuesto. En este último paradigma, la clasificación espectral resultante se compila en la figura 5.7, usando un color para cada solución subarmónica distinta, en seis pájaros.

Los distintos símbolos (colores) representan las características espectrales. rojo (+) para oscilaciones de período 1 de gran amplitud, verde (×) para soluciones de período 2 de gran amplitud, azul (*) para soluciones de período 3 y violeta para las frecuencias de período 1 y pequeña amplitud. Cada ave analizada se ubica en el eje y , y la frecuencia fundamental f_0 en el eje x .

Notablemente, el resultado es una serie de zonas bien definidas que presentan

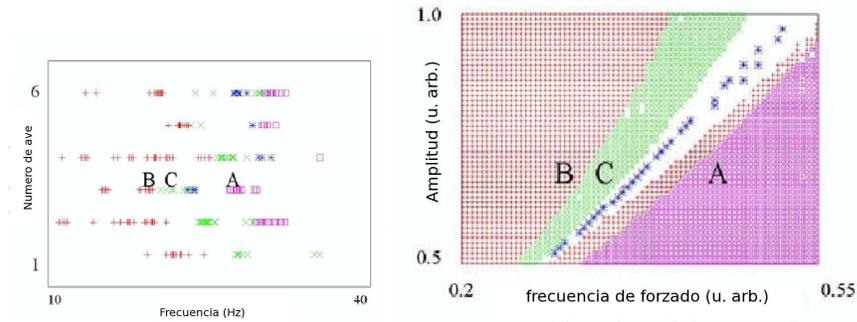


Figura 5.7: Izquierda: clasificación de patrones experimentales de presión para 6 aves según el paradigma subarmónico. Las letras A, B y C corresponden a la serie de la figura 5.6(a). Derecha: estructura de subarmónicos para el modelo de ecuaciones 5.13, 5.14 y 5.15.

la misma organización que las lenguas de Arnold del modelo (figura 5.7).

En el esquema propuesto, toda la diversidad de los patrones respiratorios durante el canto emergen como soluciones distintas de un único sistema neuronal simple que controla el sistema respiratorio. Este sistema recibe instrucciones de los núcleos superiores, que sólo deben ‘sintonizar’ la frecuencia adecuada para generar los distintos patrones de respiración.

5.3.1. Conclusiones

El modelo respiratorio propuesto en las ecuaciones 5.7, 5.8 y 5.9 permite ajustar patrones complejos para una gran variedad de aves de subgrupos lejanos en la escala evolutiva, anatómicamente distintos. Los resultados obtenidos refuerzan la idea de que la complejidad no es el resultado de un intrincado sustrato neuronal, sino de la explotación de propiedades no lineales muy genéricas de sistemas neuronales simples, que admiten un modelado de baja dimensión.

La consistencia del conjunto de resultados obtenidos en este capítulo invitan a describir los patrones vocales en los siguientes términos:

- En nuestro esquema, que no había sido explorado antes, la complejidad acústica no emerge de acciones individuales en un sistema neuronal extenso, sino de las no linealidades que describen la actividad conjunta de la red. Esta hipótesis es alternativa al consenso general según el cual hay propiedades acústicas asociadas a la acción de neuronas particulares.
- La complejidad proviene de la interacción entre los centros neuronales respiratorios y los órganos del circuito respiratorio, es decir, solamente los núcleos más bajos en la jerarquía del control motor están involucrados.

- Los patrones respiratorios siguen las mismas reglas de organización universales que los sistemas excitables forzados de baja dimensión. Las leyes que regulan la complejidad acústica pueden ponerse en términos de la organización de soluciones subarmónicas.
- Evidencia directa para el mecanismo propuesto en este trabajo implica mediciones simultáneas de la actividad de un núcleo telencefálico (por ejemplo, RA) y de un núcleo respiratorio (como RAm).

Este modelo representa los elementos esenciales de los centros respiratorios de muchas especies de aves, por lo tanto es esperable que sea ampliamente aplicable y permita explicar mecanismos vocales generales del canto de las aves.

Más aún, el modelo muestra un mecanismo simple capaz de generar diversos ritmos en general, independientemente de la arquitectura específica. Nuestros resultados permiten crear confianza en las capacidades de las reglas no lineales como fuente de complejidad en la creación de patrones neuronales.

5.4. Control neuronal y aprendizaje vocal

Los sistemas biológicos son capaces de generar una variedad extremadamente grande de comandos motores. Más aún, en muchos casos estos comandos articulados son el resultado del aprendizaje, como en el caso de los oscinos revisado en la sección precedente. Los procesos dinámicos involucrados en el aprendizaje son objeto de grandes esfuerzos de investigación tanto en cuanto al conocimiento de los mecanismos desarrollados a lo largo de la evolución biológica como respecto del diseño de sistemas artificiales capaces de adaptación y aprendizaje.

El camino de control motor para los oscinos sugiere que las neuronas premotoras en RA están gobernadas por la actividad de una subpoblación de neuronas del HVC que proyectan al RA [HKF02, McC87], figura 5.1. Además de éste, hay otro circuito neuronal muy estudiado, llamado AFP. En la figura 5.8 se esquematizan los elementos del circuito. Dos tipos de experimentos permiten relacionar este circuito con el aprendizaje. Primero, este circuito contribuye sólo mínimamente a la producción del canto en adultos [KTD03]. Además, las lesiones provocadas en este camino alteran profundamente la capacidad de desarrollar canto normal [BMA84, SchNot91].

Las neuronas de RA reciben actividad proveniente de ambos núcleos, HVC y LMAN. Esta última señal, sin embargo, requiere un tiempo de procesamiento a través de AFP (~ 40 ms en *Taeniopygia guttata*, por ejemplo).

Para modelar la dinámica de las instrucciones neuronales cíclicas y la acción del retraso en el circuito AFP, introducimos tres variables fásicas ϕ_1 , ϕ_2 y ϕ_3 ($\phi_i \in S^1$). En términos del problema modelado, ϕ_1 representa la oscilación de un núcleo como el HVC, ϕ_2 representa la actividad en alguna región del RA que contiene neuronas premotoras que controlan algún aspecto del canto, y ϕ_3 parametriza la actividad del camino indirecto, es decir que su dinámica es la

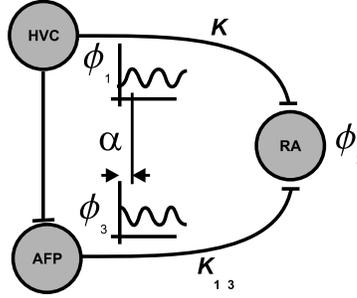


Figura 5.8: La actividad de HVC se representa con la fase ϕ_1 que gobierna las neuronas de RA a través de un camino ‘directo’ (ϕ_2) y uno indirecto (ϕ_3), que está retrasado respecto de ϕ_1 en α . La fortaleza del acople entre ϕ_2 y ϕ_1 depende de la diferencia de fase $\phi_2 - \phi_1$.

de ϕ_1 retrasada un tiempo τ . Las ecuaciones dinámicas que resultan de este escenario son

$$\dot{\phi}_1 = \omega_1, \quad (5.16)$$

$$\dot{\phi}_2 = \omega_2 - k \sin(\phi_2 - \phi_1) - k_{13} \sin(\phi_2 - \phi_3), \quad (5.17)$$

$$\dot{k} = \gamma \cos(\phi_2 - \phi_1) - k, \quad (5.18)$$

con $\phi_3 = \phi_1 - \alpha$.

El sistema puede ser reducido reemplazando $\phi_3 = \phi_1 - \alpha$ y rescaleando las ecuaciones. Las siguientes ecuaciones valen para la diferencia de fase $\phi = \phi_2 - \phi_1$ y para k

$$\dot{\phi} = 1 - k \sin(\phi) - k_{13} \sin(\phi + \alpha), \quad (5.19)$$

$$\dot{k} = \epsilon(\gamma \cos(\phi) - k). \quad (5.20)$$

Estamos interesados aquí en la dinámica de este sistema de ecuaciones, especialmente en las soluciones estacionarias, que hacemos corresponder con soluciones que el sistema puede ‘aprender’, y las ventajas que el sistema pueda sacar para distintos valores del retraso τ .

5.4.1. Soluciones e interpretación

En la figura 5.9 se muestran las nulclinas, que son las curvas que anulan cada una de las derivadas temporales para distintos valores de los parámetros. Las intersecciones de las curvas son los puntos fijos del sistema (k_0, ϕ_0) . Estas curvas son

$$k = [1 - k_{13} \sin(\phi + \alpha)] / \sin(\phi), \quad (5.21)$$

$$k = \gamma \cos(\phi) \quad (5.22)$$

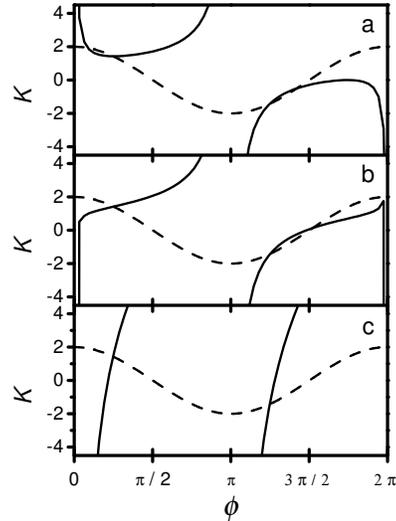


Figura 5.9: Nulclinas $\dot{\phi} = 0$ (con línea sólida), $\gamma = 2$, $\alpha = 3\pi/4$ y $k_{13} = 1,5$ (a), $k_{13} = 1,0$ (b) y $k_{13} = 10,0$ (c).

En el rango de parámetros $k_{13} \in [0, 1,42]$, la nulclina de la ecuación 5.21 presenta dos ramas, la primera con un mínimo y la segunda con un máximo. Dependiendo de los parámetros, una o ambas ramas pueden cruzar la nulclina de la ecuación 5.22. En estas intersecciones, bifurcaciones de nodo-silla ocurren. Para γ suficientemente grande, dos nodos atractores separados por dos puntos silla pueden coexistir. En el otro extremo, para γ suficientemente pequeño, no hay intersecciones y por lo tanto no se llega a una diferencia de fase estacionaria entre ambos osciladores (forzante y el forzado).

La organización topológica de las nulclinas presenta cambios cualitativos a medida que los parámetros varían. En la figura 5.10, las posiciones de los puntos fijos se grafican en función del retraso α para cuatro valores de k_{13} . Las líneas sólidas y las cruces corresponden a soluciones estables e inestables, respectivamente. Para acoplamiento suficientemente bajos, existen regiones sin soluciones estables 5.10(a), que dan lugar a la aparición de pares de puntos fijos de estabilidad opuesta en bifurcaciones de nodo-silla. En términos de las nulclinas, esta situación corresponde al mínimo de la segunda rama de la ecuación 5.21 atravesando la nulclina de la ecuación 5.22 (figura 5.9).

A medida que el acople de refuerzo aumenta, las regiones que no presentan soluciones disminuyen su tamaño hasta desaparecer en un punto transcrito (figura 5.10)(c) y (d). La existencia de estas bifurcaciones tiene un profundo

impacto en los términos de aprendizaje como los comprendemos aquí. Notablemente, para ciertos valores del retraso α , cambios mínimos dan lugar a un gran rango de diferencias de fase ‘aprendidas’ por el sistema.

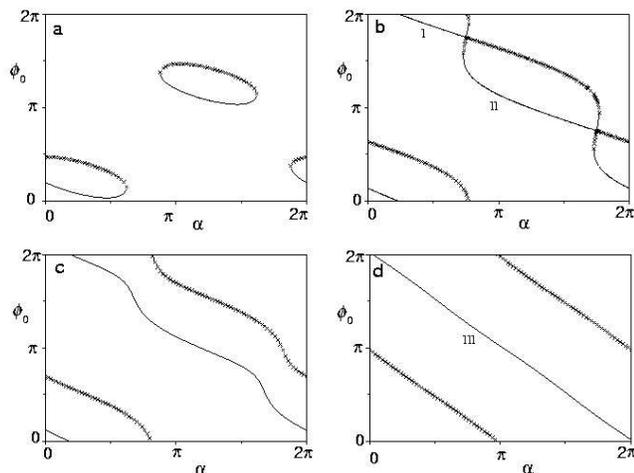


Figura 5.10: Soluciones estacionarias para $\phi(\alpha)$ con $\gamma = 1$. Los resultados para $k_{13}=0.9, 1.5, 1.8$ y 15 se muestran en los paneles (a), (b), (c) y (d) respectivamente. Las líneas sólidas corresponden a las soluciones estables, y las cruces indican ramas de soluciones inestables.

Según el modelo presentado para la dinámica de la siringe 5.2.1, la repetición de dos gestos fisiológicos esenciales son para la generación de sonido por el órgano vocal: la presión de los sacos aéreos $p(t)$ y la tensión de los labios $k(t)$ (ecuación 5.5). Más aún, la diferencia de fase entre estos gestos contiene gran cantidad de información acústica de la identidad silábica [MinGar03], parte de la cual se origina en RA. Existen conexiones directas entre nXIIIts y RA que pueden afectar la diferencia de fase entre los gestos, pero es siempre en RA donde las neuronas que reciben conexión con HVC también reciben actividad del camino AFP, y por consiguiente es este nivel donde las diferencias de fase entre gestos puede ser alterada.

Para aplicar estos resultados a un modelo de aprendizaje, asumimos que la actividad en RA está guiada por dos osciladores que representan su actividad cíclica durante la producción del canto. Uno de los osciladores representa la población de neuronas que controla la actividad respiratoria, y el otro la actividad de la subpoblación que enerva el núcleo nXIIIts. Ambos, a su vez, están controlados por la actividad del HVC, que presenta una actividad global cíclica (ritmo silábico).

El oscilador de refuerzo controla a ambos osciladores de RA con una señal idéntica a la de HVC, retrasada en una fase α . El esquema propuesto se grafica

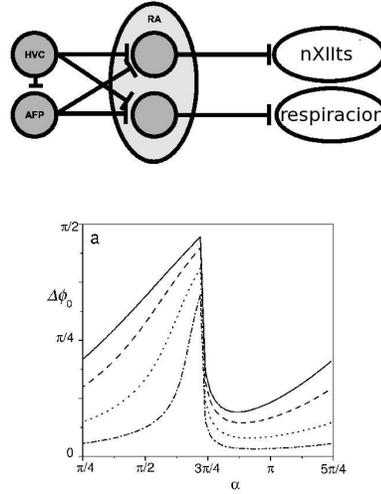


Figura 5.11: (a) Dos osciladores representando la actividad cíclica de poblaciones neuronales controladas por un oscilador principal (HVC), al que seguirán con distintas diferencias de fase. Los osciladores controlan distintos aspectos de la producción vocal: el núcleo siríngeo y los músculos respiratorios. (b) La diferencia de fase $\Delta\phi_0$ como función del retraso α . La línea sólida representa la diferencia de fase para $k_{13} = 15$ y $k_{13} = 1,5$. La línea segmentada, $k_{13} = 5$ y $k_{13} = 1,5$, la línea punteada $k_{13} = 2,5$ y $k_{13} = 1,5$ y la línea mixta $k_{13} = 1,8$ y $k_{13} = 1,5$. Para todas las simulaciones, $\gamma = 1$.

en la figura 5.11.

Si los dos osciladores están fuertemente reforzados por AFP (k_{13} es grande como en la figura 5.10) habrá un estado ‘ligado’ para cualquier valor de α . La situación es muy distinta si uno de los osciladores presenta un acople k_{13} distinto. Si uno de ellos corresponde al valor usado en la figura 5.10b y el otro al valor de la figura 5.10d, entonces, dependiendo del retraso α se aprenderán diferencias de fase cualitativamente distintas entre cada oscilador en RA.

En la figura 5.12, se muestra la diferencia de fase $\Delta\phi_0$ en función de α para distintos pares de osciladores ($\gamma = 1$). En todos los casos uno de los acoples está fijo en $k_{13} = 1,5$. Notablemente, hay valores del retraso (ceranos al valor $\alpha = 3\pi/4$ para los parámetros usados) para los cuales se puede provocar un cambio enorme en el valor estacionario de la diferencia de fase entre los osciladores. Este resultado, además, es robusto con respecto a cambios en la variable γ .

La figura 5.12 ilustra el aprendizaje de sílabas distintas para pequeños cambios en el retraso con $\alpha \sim 3\pi/4$. Los sonogramas muestran la evolución de la frecuencia fundamental del sonido producido por el modelo de siringe (ecuación 5.5) usando las diferencias de fase ‘aprendidas’.

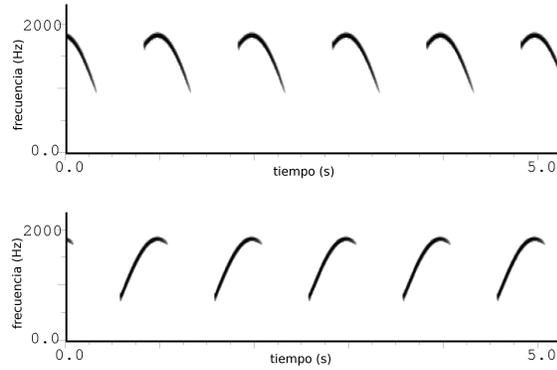


Figura 5.12: Sonogramas para dos tipos de sílaba generadas aprendiendo una diferencia de fase de (a) $\Delta\phi_0 = 0$ y (b) $\Delta\phi_0 = \pi/2$. Los parámetros usados para el modelo de siringe (ecuación 5.5) son $k(t) = 7 \times 10^7 + 6 \times 10^7 \cos(2\pi/44100)$ y $p(t) = 5 \times 10^2 + 10^3 \cos(2\pi/44100 + \Delta\phi_0)$, en unidades arbitrarias.

5.4.2. Aplicación a modelos de actividad promedio

El modelo de Kuramoto para la evolución de la diferencia de fase entre osciladores acoplados constituye un modelo muy difundido, particularmente útil para descripciones teóricas. En esta sección reproducimos este esquema para los modelos de actividad neuronal media usados a lo largo del capítulo.

Para eso, usamos las variables x e y para representar las actividades de las subpoblaciones excitatorias e inhibitorias respectivamente. En las porciones de RA cuyas neuronas proyectan al nXIIIs y a los núcleos respiratorios, las neuronas proyectoras excitatorias coexisten con otras inhibitorias. Las ecuaciones que representan su dinámica son gobernadas por dos señales. Una representa la del núcleo HVc y la otra la del núcleo LMAN, que se asume como una copia retrasada del original. Aquí usamos una típica regla hebbiana para describir el acople entre HVc y RA. El sistema resulta

$$\dot{x} = -x + S(\rho_x + ax + by + k\cos(\omega t) + k_{13} \cos(\omega t - \alpha)), \quad (5.23)$$

$$\dot{y} = -y + S(\rho_y + cx + dy), \quad (5.24)$$

$$\dot{k} = \lambda x \cos(\omega t) - k. \quad (5.25)$$

Integrando estas ecuaciones para el rango completo de retrasos $\alpha \in [0, 2\pi)$ hay algunas regiones, como la correspondiente a la figura 5.13 alrededor de $\alpha = 29\pi/15$, donde dos regiones cualitativamente distintas de período uno están separadas por una burbuja de período dos en el espacio de parámetros. Las soluciones de período uno siguen al forzante a diferentes fases. A medida que k_{13} se incrementa, sólo una solución de período uno existe. De aquí se desprende que dos subpoblaciones reforzadas por un retraso α alrededor de este valor de

bifurcación se enganchará al forzante en un amplio rango de diferencias de fase con un mínimo cambio en el retraso.

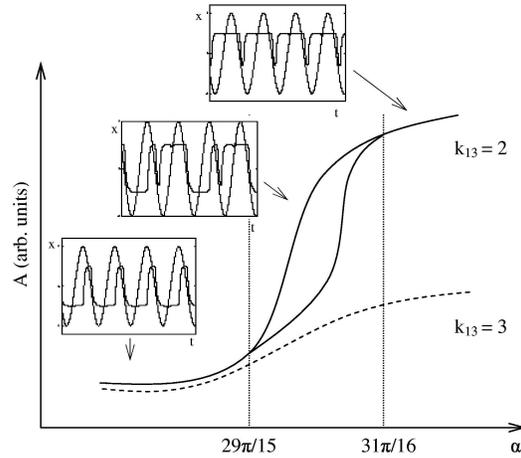


Figura 5.13: Diagrama de bifurcaciones para el sistema tipo Wilson-Cowan (ecuaciones 5.23, 5.24 y 5.25) forzado con frecuencia $\omega = 0,3$. La línea horizontal representa el retraso α y el vertical representa la máxima amplitud A de las oscilaciones en el espacio de fases. En la angosta franja $29\pi/15 < \alpha < 31\pi/16$ los osciladores con $k_{13} = 2$ y $k_{13} = 3$ logran una gran diferencia de fase entre sí. En los paneles se ven las soluciones $x(t)$ y el forzante $\cos(\omega t)$ integradas con los siguientes parámetros: $\rho_x = -5,75$, $\rho_y = -1$, $a = 10$, $b = -1,5$, $c = 2$, $d = 2$ y $\lambda = 68$.

En un trabajo reciente [SYT02] se describió el proceso de aprendizaje de una diferencia de fase entre osciladores acoplados a través de un coeficiente de acople lentamente variable. Aquí, en cambio, tratamos las consecuencias de someter el sistema de dos copias retrasadas a una señal. Su diseño se inspira en el problema del aprendizaje vocal en los oscinos, capaces de aprender complejas estructuras vocales cambiando la arquitectura de conexiones del camino de control motor. La caricatura del sistema animal analizado aquí revela una dependencia fuerte de la diferencia de fase aprendida con el retraso de la señal, lo que sugiere estrategias simples para el aprendizaje de gestos motores a través de la ‘sintonización’ de distintas subpoblaciones neuronales, logrando a un tiempo una gran plasticidad para el tipo de sílaba resultante con cambios mínimos en el retraso del refuerzo.

El control del tiempo en el camino indirecto AFP puede modificarse a través del uso de dopamina [ATMRG04], pero poco se sabe acerca de la naturaleza de la codificación usada por las aves para traducir un ‘error’ vocal en la señal indirecta. Aún no hay medidas en LMAN de aves juveniles mientras aprenden su canto. El modelo aquí desarrollado permite una exploración teórica acerca

de los mecanismos en el esquema de aprendizaje, compatibles con el modelo animal.

5.5. Discusión

En este capítulo se apuntaron ciertas direcciones para profundizar un estudio del control neuronal vocal humano. El caso del control motor en aves presentado articula directamente con algunos aspectos de aquel, mientras que otros quedan simplemente esbozados. El control neuronal de la frecuencia fundamental en el habla, por ejemplo, parece poder ser tratado como una extensión del caso presentado en este capítulo, mientras que el rol del sistema respiratorio durante las vocalizaciones presenta divergencias.

Sin embargo, es menos en las semejanzas morfológicas y anatómicas que este capítulo debe ser entendido, creemos, al proponerlo como una introducción al control motor vocal, que como un paso en la dirección del punto de vista elegido para su desarrollo. Un problema central de la neurociencia es, de hecho, la pertinencia de un nombre propio para describir el conjunto de principios y resultados que provienen de la variedad de disciplinas que la construyen. En este sentido, la búsqueda de una escala de descripción adecuada para el conjunto de problemas relacionados con el sistema nervioso es un asunto casi fundacional. La capacidad de adaptación y decisión son procesos que emergen de enormes redes de neuronas densamente interconectadas. La versatilidad de estas redes, su enorme capacidad de adaptación y la diversidad de comportamiento que producen puede sugerir que una escala para su descripción sea el del sistema extenso, con todos sus grados de libertad.

En este trabajo se explora otro punto de vista, que explota las propiedades conjuntas de las neuronas condensadas en regiones acotadas, los núcleos. Cada neurona es una unidad no lineal de procesamiento cuya actividad puede ser representada por un sistema de ecuaciones no lineales de baja dimensión. La dimensionalidad del conjunto de neuronas individuales escala con su cantidad, y su conectividad, con el cuadrado de esa cantidad.

Sin embargo, la *actividad neuronal media* puede ser representada por un sistema de baja dimensión. Los resultados presentados sugieren que la pérdida de grados de libertad no compromete la diversidad de soluciones, que en este paradigma está asociada a la no linealidad del sistema. En este capítulo hemos tratado bajo este punto de vista el problema de la complejidad acústica del canto de las aves, con resultados consistentes con una gran variedad de arquitecturas neuronales distintas.

En su diversidad de puntos de vista, esta tesis está recorrida por un espíritu de síntesis, de búsqueda de los mecanismos fundamentales, físicos y neuronales, de la producción vocal. Los modelos propuestos y analizados en este trabajo pueden leerse bajo esa luz: todos los modelos desarrollados son de baja dimensión y están inspirados y contruidos en términos de un puñado de variables biológicas fácilmente controlables. Esto permite el desarrollo de circuitos electrónicos

biomiméticos, como la siringe electrónica construida recientemente [ZysMen05].

Más allá del posible reemplazo de órganos o partes ejecutoras biológicas por artefactos biomiméticos, los resultados obtenidos aquí permiten especular con la posibilidad de desarrollar tecnologías neuromiméticas, capaces de interactuar con sustratos neuronales biológicos [ABMM03] para poder suplantar núcleos neuronales completos a través de circuitos electrónicos simples.

Apéndice A

Algoritmos genéticos

A.0.1. Adaptación del algoritmo a un problema matemático simple

En general, la adaptación de un algoritmo genético a un problema de parámetros x_i se reduce a encontrar funciones escalares $f(x_1, x_2, \dots, x_n)$ proporcionales a la similitud de la solución encontrada con la verdadera solución del problema. Esta función, en tanto explícita, permitiría resolver el problema analíticamente. Sin embargo, en los problemas de interés la función es desconocida y no hay métodos generales para encontrarla.

Por simplicidad supongamos un problema bidimensional para el que cada individuo de una población queda definido por un par de parámetros $(x; y) \in [a, b] \times [c, d]$.

Para definir la ‘codificación genética’ del individuo (x, y) , definimos un cromosoma como el arreglo consecutivo de dos genes, uno para cada número del par (x, y) . Este arreglo se construye normalizando cada coordenada por su dominio, guardando los primeros n decimales. Por ejemplo, para el par $(0,5; 1,34) \in \{[0, 1] \times [0, 2,35]\}$, la normalización arroja el par $(0,5; 0,57021276\dots)$. La identificación del individuo con su cromosoma resulta, usando cuatro cifras significativas $(0,5; 1,34) \rightarrow [50005702]$. La longitud del cromosoma es, en este caso, 8.

En el ‘espacio genético’ así definido se pueden aplicar los operadores de cruzamiento y mutación, que en la evolución biológica suceden en el espacio de las bases nitrogenadas y, en este ejemplo, en la base decimal. A una mutación corresponderá el reemplazo de cualquiera de los 8 números del cromosoma por otro, por ejemplo, $[23126675] \rightarrow [23026675]$. El cruzamiento consiste en el intercambio, a partir de cualquier posición, de la información de los cromosomas de los individuos seleccionados. Por ejemplo, $[23126675] + [87654321] \rightarrow [23154321]$.

Con estas definiciones, el algoritmo genético queda adaptado a un problema genérico bidimensional, y su ejecución consiste en elegir una población inicial de N pares (x_i, y_i) , seleccionarlos según su adaptación usando una función $f(x_i, y_i)$ y aplicarles los operadores genéticos para generar una nueva población. El proce-

so se repite entonces para un número fijo de generaciones, o hasta que satisfacer alguna condición buscada.

A.0.2. Bases teóricas de los algoritmos genéticos

Definiciones

Por simplicidad, asumimos que la codificación genética de un individuo, es decir su *genotipo*, puede expresarse por un arreglo binario de longitud L al que llamaremos cromosoma. Por ejemplo, para $L = 10$,

$$c = |1011011001| \quad (\text{A.1})$$

representa uno de los cromosomas posibles del conjunto designado por $[1, 0]^L$, que tiene $M = 2^L$ elementos.

Esta representación permite, por ejemplo, una analogía directa con la codificación biológica. Asignando un número a cada base nitrogenada del ADN, A (0), T (1), C (2) y G (3), los pares binarios de la ec. A.1 representan la secuencia genética *CGTCT*. El genoma humano, por ejemplo, tiene alrededor de tres mil millones de pares de bases. Como comparación de las dimensiones del problema, la cantidad de cromosomas binarios distintos de longitud $L = 100$ es de $M_{bin} = 2^{100} \sim 10^{30}$.

En el caso del modelo de voz, la elección de 4 cifras decimales significativas para cada uno de los 9 parámetros de la fuente y del filtro, representa una cantidad $M_{voz} = 10^{36}$ de cromosomas distintos. Un dado cromosoma de voz se representará por $c_v = |m, k, b, R_2, I_2, A_1/A_2, A_2/A_3, L_1, L_2|$. Cada posición (*gen*) se define como las primeras 4 cifras significativas del desarrollo decimal de los parámetros A_1, A_2, A_3, L_1 y L_2 del tracto vocal y m, k, b de la fuente (sección 2.2.2), normalizadas.

El concepto principal en el desarrollo de los algoritmos genéticos es el de *bloque constructivo*, que define un subconjunto de genes dentro del cromosoma. Por ejemplo,

$$b = |1011\star\star\star\star 0\star|. \quad (\text{A.2})$$

En este caso, el símbolo \star representa una posición que no será tenida en cuenta en la evaluación del cromosoma. El conjunto de los bloques constructivos binarios está determinado por $[1, 0, \star]^L$. En el caso de la voz, por ejemplo $|\star\star\star\star A_1/A_2 A_2/A_3 L_1 L_2|$, representa el bloque constructivo de todas las soluciones que mantienen fija la estructura del tracto vocal, independientemente de la fuente.

La idea subyacente de los algoritmos genéticos es su capacidad de preservación de los mejores bloques constructivos y su combinación para generar individuos mejor adaptados. A lo largo de las generaciones se producen producen los siguientes pasos:

1. Reproducción de acuerdo a la adaptación: se seleccionan cromosomas con una probabilidad proporcional a su adaptación.
2. Recombinación: los operadores de crossover y mutación generan los nuevos individuos.

3. Reemplazo: los individuos de la nueva población toman el lugar de los padres para recomenzar el proceso.

Sin embargo, tanto el crossover como la mutación tienden a romper bloques constructivos en las sucesivas generaciones. En lo que sigue se tratará de entender el rol de ruptura de los operadores genéticos en la búsqueda de soluciones óptimas.

Reproducción proporcional a la adaptación

Para estudiar la evolución de los bloques constructivos a lo largo de las generaciones, el primer problema que surge es el de estimar el valor $S(b)$ de un bloque constructivo b . Una posibilidad es la de calcular el promedio S de adaptación de todos los cromosomas, y luego el promedio de los compatibles con el bloque b . El bloque b es mejor que el promedio si $S(b) > S$.

En general, este cálculo es imposible; por ejemplo, el reemplazo de cualquier posición en el cromosoma c de la ec. A.1 por una \star , creará un bloque constructivo que lo incluya. Podemos dejar invariante cada posición del cromosoma o cambiarla por \star . Esto arroja 2^{100} posibles bloques constructivos compatibles con cada una de las reglas posibles.

La primera pregunta entonces es cómo evolucionan los bloques constructivos bajo el proceso de selección. Para la siguiente descripción, se asume por simplicidad que la adaptación de un cierto cromosoma determina el número de descendientes que producirá en la siguiente generación, y que un individuo promedio produce 1 nuevo individuo.

Supongamos entonces que el bloque $b_1 = |1\star\star\dots\star|$ tiene 3 instancias en una población, con adaptaciones de 1, 0 y 1 respectivamente. Las tres instancias de b_1 producirán $1+0+1=2$ descendientes, un promedio de $2/3$ por cada instancia. Este número es menor que el del individuo promedio, y el bloque $|1\star\star\dots\star|$, estará presente 2 veces en la próxima generación.

Ahora supongamos que $b_2 = |\star 0 \star 1 1 \star \dots \star|$ tiene también tres instancias en una población, con adaptaciones 2, 2 y 1 respectivamente. El promedio será esta vez de $5/3$ por instancia, que es mayor que el promedio, y por lo tanto la nueva generación presentará más instancias del bloque b_2 .

En resumen, a partir de cualquier bloque $b \in [1, 0]^L$, si $M(b, t)$ es el número de instancias de b en la generación t , entonces

$$M(b, t+1) = S(b, t)M(b, t) \tag{A.3}$$

da el número de instancias en la nueva generación, $t+1$, después de la reproducción. Sin embargo, este resultado no es tan bueno como aparece a simple vista, porque a pesar de seleccionar los mejores bloques, genera más y más copias idénticas en las nuevas generaciones. Sin importar cuán grande la población inicial, se trata de una parte minúscula de las posibilidades. El crossover aparece entonces como una posible solución.

Efectos del crossover

Este proceso es capaz de recombinar distintos bloques sin perturbar el resultado obtenido antes, es decir, permitiendo que los bloques con mayor adaptación se preserven y aumenten sus instancias en las sucesivas generaciones de la población.

El efecto del crossover depende de la longitud $L(b)$ del bloque b , definida como el número de sitios de crossover posibles entre los sitios extremos que definen el bloque. Por ejemplo, en el bloque $|\star 0 \star 1 1 \star \dots \star|$, las posiciones 2 y 5 definen sus extremos. Hay tres posibles sitios de crossover que podrían romper el bloque, de manera que $L(|\star 0 \star 1 1 \star \dots \star|) = 3$.

Los bloques más cortos serán menos sensibles a la ruptura por crossover. Los que sobrevivan pasarán a la próxima generación según la prescripción de A.3. En un cromosoma de longitud L hay $L - 1$ sitios posibles para crossover, de manera que la probabilidad de que el crossover *no* caiga en los límites del bloque es $1 - L(b)/(L - 1)$. Entonces, la corrección a la ecuación A.3 por los efectos de crossover será

$$\langle M(b, t + 1) \rangle = \left[1 - \frac{L(b)}{L - 1} \right] S(b, t) \langle M(b, t) \rangle. \quad (\text{A.4})$$

Hay dos argumentos por los que la tendencia a la ruptura de los bloques de mayor longitud no ofrece mayores problemas. Primero, este proceso tiende a descubrir rápidamente los bloques más pequeños. La gran cantidad de instancias de bloques pequeños en la población permite la estimación de una enorme cantidad de bloques (vale recordar que la cantidad de bloques presentes en un dado cromosoma es de aproximadamente 10^{30}).

Luego, se observa que los bloques de mayor longitud se componen de combinaciones de bloques menores, de suerte que se establece una composición jerárquica en la que bloques de un nivel se combinan para construir bloques del siguiente.

Efectos de la mutación

El rol de las mutaciones, en el esquema presentado, está relacionado con los efectos de saturación de poblaciones finitas. Es posible, por ejemplo, que algún bloque muy simple termine presente en toda la población. Este bloque, bajo el efecto de la reproducción y el crossover, no será removido nunca. En particular, si fuera el caso para el bloque $|1\star\dots\star|$ la exploración de la población se reduciría a la mitad. Para evitar este efecto de *fijación*, la mutación se propone como la probabilidad P_{mut} de cambiar un sitio del cromosoma al azar. El número de instancias de cada bloque b , incluyendo el efecto de las mutaciones, resulta

$$\langle M(b, t + 1) \rangle = \left[1 - \frac{L(b)}{L - 1} \right] [1 - P_{mut}(b)] S(b, t) \langle M(b, t) \rangle. \quad (\text{A.5})$$

Esta fórmula es conocida como el *teorema del bloque constructivo* para algoritmos genéticos.

El número de bloques constructivos es tan grande que no resulta posible calcular explícitamente una estimación para su adaptación. Los algoritmos genéticos hacen este cálculo *implícitamente*, operando sobre una población finita de cromosomas pero actuando como si pudieran hacer estimaciones sobre los bloques. Esta habilidad para manipulaciones *implícitas* de grandes números a través de cálculos *explícitos* sobre poblaciones relativamente pequeñas es conocido como *paralelismo implícito*. La evolución, así concebida, permite ‘recordar’ combinaciones de bloques que incrementan la adaptación, aun ‘olvidando’ individuos particularmente adaptados.

A.1. Paralelización del algoritmo

La idea básica de la implementación en paralelo consiste en elegir un procesador, el *nodo principal*, para separar la población inicial de cromosomas en sub-dominios a ser procesados en paralelo por el resto de los procesadores (*nodos esclavos*). Cada uno de los nodos esclavos encuentra la adaptación de cada subpoblación, y esta información es devuelta al nodo principal, que aplica los operadores genéticos a la población completa, generando una nueva población que repetirá el proceso.

En la figura A.1 se grafica la evolución del tiempo de ejecución del código con la cantidad de procesadores.

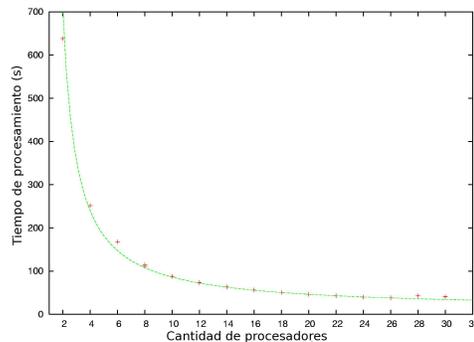


Figura A.1: Tiempo de ejecución vs. cantidad de procesadores para el algoritmo genético usando 500 generaciones, 150 cromosomas, tasa de crossing del 40% y tasa de mutaciones de 8%. Los valores usados resultan del promedio de 2 o 3 corridas. A partir de $n = 10$ procesadores, la dispersión en los valores es despreciable.

A primer orden, se espera que el tiempo de procesamiento T_p en función del número de procesadores n se comporte aproximadamente como $T_p(n) \propto 1/n$, mientras que el tiempo de comunicación $T_c(n) \propto n$. Los ajuste devuelven un valor despreciable para la corrección lineal α (ver figura A.1), de modo que para la cantidad de procesadores usados, la red de procesadores está lejos de saturar. Estas simulaciones se llevaron a cabo en el cluster de la FCEyN.

Apéndice B

Métodos

B.1. Medición de la presión de sacos aéreos

La presión de los sacos aéreos fue registrada en las aves anestesiadas mediante la inserción una cánula a través de la pared abdominal, posterior a la última costilla, de manera que penetre unos milímetros en el saco aéreo torácico.

El extremo libre de la cánula se conecta a un transductor de presión piezo-resistivo montado en el dorso del ave. Desde allí las señales se transmiten a un amplificador, desde donde, después de multiplicarlas por una onda sinusoidal de 1kHz, se pueden registrar usando una PC con tarjeta de sonido. Allí las señales fueron digitalmente demoduladas. [TMM06]

Bibliografía

- [ABMM03] J. Aliaga, N. Busca, V. Minces, G.B. Mindlin, B. Pando, A. Salles, L. Szczupak *Phys. Rev. E* **67**, 061915, 2003.
- [Aba96] H.D.I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, 1996.
- [ATM05] A. Amador, M.A. Trevisan, G.B. Mindlin, *Phys. Rev. E* **72**, 031905, 2005.
- [ATMRG04] H.D.I. Abarbanel, S. Talathi, G. B. Mindlin, M. Rabinovich, L. Gibb, *Phys. Rev. E*, **70**, 051911, 2004.
- [BraDou02] M.S. Brainard, A.J. Doupe. *Nature* **417**,351-358.
- [BMA84] S. Bottjer, E.A. Miesner, A.P. Arnold, *Science* **224**, 901, 1984.
- [McCKon81] J. S. McCasland, M. Konishi, *Proc. Nat. Acad. Sci.* **78**: 7815-7819, 1981.
- [McC87] J.S. Mc Casland *et al.* *J. Neurosci.* **7**, 2339 (1987).
- [Dub98] J.L. Dubbeldam, *Neth. J. Zool.* **48**:3, 199-212, 1998.
- [Ele03] C. P. H. Elemans, O. N. Larsen, M. R. Hoffmann, J. L. van Leeuwen. *Anim. Biol.* **53**, 183-193, 2003.
- [Fee99] M. S. Fee, B. Shraiman, B. Peseran and P. P. Mitra, *Nature (London)* **395**, 67, 1998.
- [FGPV88] M. Feingold, D. L. Gonzalez, O. Piro, H. Viturro, *Phys. Rev. A* **37**, 10, 1988.
- [Fla72] J. L. Flanagan, *Speech Analysis: Synthesis and Perception* Springer-Verlag, New York, 1972.
- [Gar01] T. Gardner, G. Cecchi, M. Magnasco, R. Laje and G. B. Mindlin. Simple motor gestures for birdsongs. *Phys. Rev. Lett.* **87**, 208101, 2001.
- [Gil98] R. Gilmore, *Topological analysis of chaotic dynamical systems*, *Rev. Mod. Phys.*, **70**4, p. 1455 1998.

- [Gil02] Gilmore, R. and Lefranc, M. *The topology of chaos*, Wiley, 2002.
- [GolSut96] F. Goller, R.A. Suthers. *J. Neurophysiol.* **75**:867-876 (1196).
- [GolSut96b] F. Goller, R.A. Suthers **76**:287-300, 1996.
- [Gol89] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, 1989.
- [GucHol83] J. Guckenheimer, P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springer, New York, 1983.
- [HKF02] R.H.R. Hahnloser, A. A. Kozhevnikov, M.S. Fee, *Nature (London)* **419**, 65, 2002.
- [Her93] H. Herzl, *Appl. Mech. Rev.* **46**, 399, 1993.
- [Hol96] Holland, *Hidden Order: how adaption builds complexity*, Addison-Wesley, 1996.
- [HopIzh97] F. C. Hoppensteadt, E. M. Izhikevich, *Weakly Connected Neural Networks*, Springer, New York, 1997.
- [IshFla72] K. Ishizaka, J.L. Flanagan, *Bell Syst. Tech. J.* **51**, 1233, 1972.
- [JTM05] P. E. Jercog, M. A. Trevisan, G. B. Mindlin, *Physica A*, 356, 145-150, 2005.
- [Kee98] J. Keener, J. Sneyd, *Mathematical Physiology*, Springer, Germany, 1998.
- [Ken79] R. D. Kent, *Journal of Speech and Hearing Disorders*, **44**, 513-521.
- [KTD03] R.R. Kimpo, F.E. Theunissen, A.J.J. Doupe, *J. Neurosci.* **23**, 5750, 2003.
- [LajMin01] R. Laje, T. Gardner, G. B. Mindlin, *Phys. Rev. E*, **64**, 056201, 2001.
- [MLG88] J. Speech Hear. Res. **31**:556-568, 1988.
- [Lev99] T. C. Levin and M. E. Edgerton, *Sci. Am.* **70**(9), 1999.
- [Lef93] M. Lefranc, P. Glorieux, *Int. J. Bifurcation and Chaos* **3**, 643 (1993).
- [MacD01] N. MacDonald et al., *Writing Message-Passing Parallel Programs with MPI*, Course Notes, Edinburgh Parallel Computing Centre, University of Edinburgh.
- [Min90] Mindlin, G. B., Hou, X., Solari, H., Gilmore, R. and N.B. Tufillaro *Phys. Rev. Letts.* **64** 2350 (1990).
- [Min92] Mindlin, G. B. and Gilmore, R. *Physica D* **58**, 229 (1992).
- [MinGar03] G. B. Mindlin, T. Gardner, F. Goller, R. Suthers, *Phys. Rev. E*, **68**, 041908, 2003.

- [MinLaj05] G. B. Mindlin, R. Laje, *The Physics of Birdsong*, Springer, Berlin, 2005.
- [Mit98] M. Mitchell, *An introduction to genetic algorithms (Complex adaptive Systems)*, MIT press, 1998.
- [Not76] F. Nottebohm, T. M. Stokes, C. M. Leonard. Central control of song in the canary. *J. Comp. Neurol.* **165**, 457–486.
- [Not93] F. Nottebohm, *Neth. J. Zool* **43**, 193-234, 1993.
- [NR99] Press, H. W. et al., *Numerical Recipes in C: The art of scientific computing*, Cambridge University, 1999.
- [Pat03] G. B. Mindlin, M. A. Trevisan. M.Eguía, *Topological voiceprints for speaker identification*, SD2003/222/1, PCT/US04/027193, 2003.
- [PetBar52] G.E. Peterson, H.L. Barney, *Journal of the Acoustical Society of America* **24**, 175-184.
- [Rab93] L. Rabiner, B. Juang, *Fundamentals of speech recognition*, Prentice Hall, 1993.
- [Rie98] F.Rieke, D. Warland, R. deRuytervanSteveninck, W. Bialek, *Spikes, exploring the neural code*, MIT Press, 1998.
- [Rol76] Rolfsen D., *Knots and Links*, Publish or Perish, Berkeley, CA.
- [Rot81] Rothemberg, *Vocal Fold Physiology*, University of Tokyo Press, Tokyo, pp. 304-323, 1981.
- [SchNot91] C. Scharff, F. Nottebohm, *J. Neurosc.* **11**, 2896, 1991.
- [SchWag90] Schuster H.G. and Wagner P., *Biol. Cyber.*, **64** 77-82 (1990)
- [Sm02] Z. M. Smith, B. Delgutte, A. J. Oxenham, *Nature (London)*, **416** 87, 2002.
- [SNM96] H. Solari, M. Natiello, G. B. Mindlin, *Nonlinear Dynamics: a two way trip from Physics to Math*, IOP, London, 1996.
- [SolGil88] Solari H. G. and Gilmore R. *Phys. Rev. A* **37**, 3096 (1988).
- [Spi99] Spiro et. al. , 1999.
- [StoTit95] B. Story, I. Titze, *J. Acoust. Soc. Am.* **97**, 1249, 1995.
- [SutMar02] Suthers R. and Margoliash D., *Curr. Opin. Neurobiol.* **12** 684 (2002)
- [SWM03] C. B. Sturdy, J. M. Wild, R. Mooney, *Journal of Neuroscience*, **23**(3), 1072-1086, 2003.
- [SYT02] Seliger P., S.C. Young, L. Tsimring, *Phys. Rev. E* **65** 041906 (2002).

- [TBSM05] M. A. Trevisan, S. Bouzat, I. Samengo, G. B. Mindlin, Phys. Rev. E **72**, 011907, 2005.
- [TEM01] M. A. Trevisan, M. C. Eguia, G. B. Mindlin, Phys. Rev. E **63**, 026216, 2001.
- [TEM04] M. A. Trevisan, M. C. Eguia, G. B. Mindlin, Physica D, **200**, 75-80, 2004.
- [Tit88] I. R. Titze, J. Ac. Soc. Am. **83**, 1536, 1988.
- [Tit93] I. R. Titze, *Principles of Voice Production*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [TMG06] A. Trevisan, G. B. Mindlin, F. Goller, Phys. Rev. Letts. **96** 058103, 2006.
- [TMM06] M.A. Trevisan, J.M. Méndez, G.B. Mindlin, Phys. Rev. E, en prensa, 2006.
- [VMK94] E. Vu, E. Mazurek, M. Konishi, J. Neurosci. **14**: 6924-6934, 1994.
- [Wil97] J. M. Wild, Eur. J. Morphol. **35**:303-325,1997.
- [Wil97b] J. M. Wild, J. Neurobiol. **33**, 653,1997.
- [YuMar96] Yu, D. Margoliash, Science. **273** 5283. 1871-1875, 1996.
- [ZysMen05] D. Zysman, J.M. Méndez, B. Pando, J. Aliaga, F. Goller, G.B. Mindlin, Phys. Rev. E **72**, 051926, 2005.

Lista

No pensé que volvería a escribir Fiorenza en ninguna parte, nunca. Ahí está de nuevo, siempre.

Mis padres, Grazia y Alberto, y mis hermanos, Leonardo y Valeria. Minuto. Puedo nombrar solamente amigos, porque la suerte me salvó del peso de colaboradores y de jefes. Gabriel, entonces. Ojalá siempre andemos juntos por ahí.

Damián, María.

Gonza, Martín, Coto, Ñato.

Manuel, Rodrigo, Pablito, Anita. Víctor en el mar. Richi.

Estuvo, casi todo lo que duró este trabajo, Julia.

Me gusta esta lista.