

Tesis Doctoral

Empleo de coordenadas motoras para decodificar el sistema motor en el canto de aves

Sanz Perl Hernández, Yonatan

2014-12-15

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Sanz Perl Hernández, Yonatan. (2014-12-15). Empleo de coordenadas motoras para decodificar el sistema motor en el canto de aves. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Sanz Perl Hernández, Yonatan. "Empleo de coordenadas motoras para decodificar el sistema motor en el canto de aves". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2014-12-15.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

**Empleo de coordenadas motoras
para decodificar
el sistema motor en el canto de aves**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por Yonatan Sanz Perl Hernández.

Director del trabajo: Dr. Gabriel Mindlin.

Consejero de estudios: Dr. Mariano Sigman.

Lugar de trabajo: Laboratorio de Sistemas Dinámicos, Departamento de Física, FCEyN, UBA

Buenos Aires, 2014

Fecha de defensa: 15 de Diciembre 2014

Empleo de coordenadas motoras para decodificar
el sistema motor en el canto de aves.

Lic. Yonatan Sanz Perl

Buenos Aires, 2014

Resumen

El trabajo que relata esta tesis se basa en la implementación de herramientas de análisis provenientes de la física en áreas vinculadas a las neurociencias. En particular esta tesis se encarga de mostrar la validez de estas herramientas y su capacidad de ser incorporadas como nuevas técnicas de análisis para abordar problemáticas típicas en neurociencias con otro punto de vista.

Las aves canoras, modelo animal empleado en esta tesis, tienen la capacidad de aprender y generar un comportamiento complejo que es el canto. Para ello cuentan con un sofisticado órgano fonador, llamado Siringe, que es controlado mediante instrucciones provenientes del sistema nervioso. Durante la última década el Laboratorio de Sistemas Dinámicos ha realizado numerosos trabajos sobre el funcionamiento la Siringe, buscando comprender cuánto de la complejidad del canto se debe a la complejidad propia del órgano fonador y cuánto se debe a complejidad en las instrucciones. Para ello se han realizado modelos matemáticos de baja dimensión, operacionales y predictivos capaces de generar sonidos sintéticos muy similares a los del ave, permitiendo asimismo comprender cuáles son las instrucciones necesarias para la obtención de dichos sonidos.

Esta tesis se encarga en primera instancia de mostrar la pertinencia de dichos modelados para el caso del Diamante mandarín en términos de 3 observables distintos, para luego ir tras la pregunta de cómo están organizadas las instrucciones en el sistema nervioso. En términos de características acústicas se comprobó los sonidos generados con el modelo presentan la misma relación entre frecuencia fundamental y contenido espectral que los sonidos reales del ave, sonidos de baja frecuencia fundamental presentan alto contenido espectral y viceversa. Por otro lado, en términos de los gestos motores se comprobó que los gestos de presión y tensión que emplea el ave durante su canto para controlar la siringe, están fuertemente correlacionados con los parámetros del modelo que dan cuenta de la presión y la tensión. Finalmente, se confirmó la pertinencia del modelo en términos de respuesta neuronal en neuronas selectivas al canto propio en núcleos vinculados al aprendizaje y generación del canto. Para ello se realizaron experimentos de electrofisiología tanto en aves despiertas como dormidas mostrando que la respuesta neuronal presentada ante el estímulo de su propio canto grabado era muy similar a la presentada ante el estímulo del canto sintético. Vale aclarar que para lograr

ese resultado fue necesario realizar una serie de mejoras al modelo, como la inclusión de un tracto vocal superior que actuara como filtro dinámico del sonido permitiendo incorporar mejoras tímbricas canto sintético.

Finalmente con el modelado validado se abordó la pregunta de cómo codifica el cerebro la generación del canto. Para ello, se propuso realizar un cambio de coordenadas, de las acústicas (donde se venía estudiando el problema desde años) a las motoras. En estas coordenadas, se pudo observar una fuerte relación entre los disparos neuronales en el HVc (núcleo vinculado al aprendizaje y generación del canto) y los extremos de los gestos motores que se obtienen a partir del modelado del canto.

Este resultado abrió un debate en el área, ya que hasta entonces el paradigma establecido afirmaba que los disparos en HVc eran como un “reloj” que dispara equidistante sin relación con la generación del canto. Este resultado, asimismo se suma a otros resultados que aparecen en la literatura de los últimos años que muestran observaciones contradictorias con el paradigma del “reloj”. En esta tesis se plantea el debate como algo abierto aún y a la espera de la existencia de un modelo de sistema motor en el canto de ave que integre todas estas observaciones recientes.

Palabras Claves: canto de aves - gestos motoras - sistema motor.

*Title: **DECODING THE MOTOR SYSTEM ON BIRDSONG USING MOTOR COORDINATES.***

Abstract

This thesis is based on the implementation of analysis tools from physics in areas related to neuroscience. In particular, this thesis seeks to show the validity of these tools and their ability to be incorporated as new analytical techniques to approach typical problems in neuroscience from a different point of view.

Songbirds, animal model used in this thesis, have the ability to learn and generate a complex behavior which is singing. In order to sing, they have a sophisticated vocal organ called Syrinx, which is controlled by instructions from the nervous system. During the last decade the Dynamical Systems Lab has performed extensively on the Syrinx operation, seeking to understand how much of song complexity is due to the complexity of the vocal organ and how much is due to complexity in the instructions. This has been done, operational and predictive low dimensional mathematical models capable of generating very similar to those of synthetic bird sounds, allowing also understand what the instructions for obtaining these sounds are.

This thesis is responsible in the first instance to show the relevance of those modeled for the case in terms Zebra finch 3 different observables, then go after the question of how instructions are organized in the nervous system. In terms of acoustic characteristics of sounds generated was found to have the same pattern relationship between the fundamental frequency and the actual spectral content of the bird sounds, sounds of low fundamental frequency and vice versa at high spectral content. On the other hand, in terms of motor gestures it was found that the pressure and tension gestures employing the bird for its song to control the syrinx, are strongly correlated with the model parameters that account for the pressure and tension. Finally, the relevance of the model was confirmed in terms of selective neuronal response to own singing in nuclei related to song learning and generation of neurons. This electrophysiology experiments were conducted in both awake and asleep birds showing that the neural response to the stimulus presented their own recorded song was very similar to that presented to the stim-

ulus of synthetic singing. It is clear that to achieve this result it was necessary to make a number of improvements to the model, the inclusion of an upper vocal tract act as dynamic filter allowing incorporate timbral sound improvements synthetic singing.

Finally modeling validated with the question of how encoding celebrated singing generation board. To do this, it was proposed to make a change of coordinates, acoustic (where the problem had been studying for years) to the motor. In these coordinates, we observed a strong relationship between neuronal firing in HVC (nucleus linked to learning and singing generation) and the ends of the motor gestures obtained from the modeling of singing.

This result opened a debate in the area, because until then the established paradigm claimed that the shots were HVc how a “ clock ’equidistant firing unrelated to the generation of singing. This result also adds to other results reported in the literature in recent years that show contradictory observations with the paradigm of “ clock ’. In this thesis the discussion as open still and waiting for the existence of a power model in the singing bird that integrates all these recent observations system arises.

Keywords: Songbird - Motor gestures - Motor system

Agradecimientos

Tengo tanta gente querida e importante para agradecerle en todo este recorrido que de hacerla completa esta sección sería realmente larga. Entonces no voy a hacer una gran lista con palabras entre emotivas y cursis, llenas de motivos para agradecer a cada quien y demás. Pero si hay una lista corta de algunas personas que quiero agradecer especialmente: antes que nada a Gabo, por su infinita generosidad y empeño en la enorme tarea de dirigir y transmitir lo que se trata hacer ciencia. Sin su guía todo esto hubiera sido imposible y realmente no hay un conjunto de palabras preciso para agradecer lo suficiente y describir lo grato y enriquecedor que fue trabajar bajo su dirección. A Ana, de quien aprendí muchísimo y disfruté trabajar con ella todos estos años. Mucha admiración por su vocación, entrega e increíble predisposición para las cosas. Al resto del LSD (actual y pasado) con quienes compartí años muy divertidos de aprendizaje y crecimiento, en un clima de trabajo sin igual, de donde me llevo un lindo paquete de amigos. A mi vieja y Camilo, que sin su apoyo y aguante hubiera tirado todo por el aire varias veces. A Clarita que me acompañó en este último tramo.

Sin estas personas que en muchísimos momentos del recorrido me han dado una mano en todos los aspectos posibles (de la ciencia y de la vida), esta tesis no hubiera sido posible en absoluto.

Entonces:

GRACIAS,

MUCHAS GRACIAS.

Índice general

1. Introducción	1
1.1. La importancia del canto de aves	2
1.2. El sistema del canto	3
1.2.1. Descripción neuroanatómica	4
1.2.2. Interpretación neuroetológica	6
1.3. Una biomecánica descriptiva	13
1.3.1. Generación de sonido.	13
1.3.2. El canto del Diamante Mandarín	14
1.3.3. Morfología de la Siringe.	16
1.3.4. Los músculos involucrados	20
1.3.5. El tracto vocal superior	22
2. Una biomecánica cuantitativa	25
2.1. El modelo físico	26
2.1.1. El modelo	26
2.1.2. Estudio dinámico	30
2.1.3. Cuantificación de características acústicas determinadas por la biomecánica.	33
2.2. La forma normal	35
2.3. Estudio e implementación de la forma normal	38
2.3.1. Ajuste factor de escala temporal	38
2.3.2. Curvas de iso-nivel para la frecuencia fundamental y el SCI	40
2.3.3. Reconstrucción de gestos motores	42

3. Una síntesis realista	47
3.1. Modelos de trato vocal superior	48
3.1.1. Modelos de tubos	48
3.1.2. Modelo con OEC	50
3.2. Pertinencia de los modelos en términos de respuesta neuronal . . .	55
3.2.1. Selectividad neuronal	55
3.2.2. Fenómeno de réplica	57
3.2.3. Comprobación de los modelos	58
3.3. Sintonización de respuesta neuronal a variaciones en los parámetros estáticos	65
4. Codificación neuronal interpretada en coordenadas motoras.	69
4.1. Los gestos motores como unidad del canto	69
4.2. Las neuronas proyectoras disparan en los extremos de los gestos motores (GTE)	71
4.2.1. Descripción del resultado, aparición de GTE como instan- cias codificantes	71
4.2.2. Estadística sobre los resultados de HVc_p	75
4.3. Las interneuronas son inhibidas en los GTEs	77
4.3.1. Procesamiento de los datos de actividad tónica	77
4.3.2. Estadística sobre los resultados de HVc_i	78
4.4. La representación de los gestos durante el canto	81
4.5. Implicancias de la codificación motora.	83
5. Automatización del algoritmo	87
5.1. Obtención de la frecuencia fundamental	88
5.2. Reconstrucción de los gestos y síntesis de canto automática	91
5.3. Selectividad en HVC	93
5.4. Empleo de la reconstrucción y síntesis automatizada para estudiar la diferencia del contexto social de emisión del canto	96
6. Conclusiones	99
Appendices	105

A. Figuras Suplementarias al Capítulo 4	107
B. Código en C	113
Bibliografía	121

a G, A.A. y la ciencia con flota-flota

Capítulo 1

Introducción

Cuando a alguno de los integrantes del Laboratorio de Sistemas Dinámicos nos preguntan de pasada sobre qué trabajamos generalmente nos encontramos con diversas expresiones de sorpresa al dar la respuesta de “*Canto de aves*“. En ese momento el interlocutor dada la brevedad de la respuesta, sigue preguntando sin salir del asombro: *Por qué estudian el canto de las aves?*, cuya respuesta breve de nuevo sería: *”las aves (algunas de ellas) son un excelente modelo animal para estudiar aprendizaje de comportamientos complejos y es ampliamente estudiado dentro de la neurociencia.* No conforme (y con razón) suelen continuar las preguntas: *Tiene algo que ver con la música?*, *Estudian cómo se comunican los pajaritos?*, *Estudian cómo hacen para cantar?*, *Se preguntan qué piensan cuando cantan?* A lo que debemos responder, nuevamente de forma breve, un poco que si a cada una de estas preguntas.

La propuesta de esta tesis es dejar de lado el laconismo de este diálogo de pasillo, mostrar cómo ese tipo preguntas están interconectadas entre sí y que encontrar alguna de sus respuestas nos ayudan a comprender los mecanismos neuronales involucrados en el aprendizaje y la generación de un comportamiento complejo. En esa dirección, las siguientes hojas son un relato del recorrido de cómo abordamos algunas de estas preguntas en el transcurso de mis años de doctorado en el Laboratorio de Sistemas Dinámicos desarrollando las herramientas necesarias para obtener algunas respuestas.

1.1. La importancia del canto de aves

El estudio del canto de aves representa una aproximación a la comprensión del funcionamiento de algunas tareas que desarrolla el cerebro, en donde el conocimiento del comportamiento resulta una guía en la búsqueda de los mecanismos neuronales subyacentes. En ese sentido, los primeros trabajos que iniciaron el estudio científico de forma sistemática en canto de aves fueron los de Thorpe [Thorpe, 1961] y Marler [Marler, 1970]. Ellos mostraron que las aves juveniles necesitan, en una primera etapa de aprendizaje, la presencia de un tutor de la misma especie para aprender su canto. Asimismo si el juvenil es aislado del tutor, el canto adulto resultante se encuentra muy deteriorado. Luego trabajos de Konishi [Konishi, 1965] mostraron que las aves juveniles necesitan a su vez escuchar su canto propio durante una etapa del aprendizaje posterior para lograr un canto típico de su especie. De esta forma se establecieron dos etapas de aprendizaje para las vocalizaciones, la *Sensorial* y la *Sensorimotora*, una sucede a la otra y en varios casos se superponen un período de tiempo. En la primer etapa, el ave crea una representación interna de lo que escucha de sus tutores, conocido como el templado, mientras que en la segunda, compara sus propias vocalizaciones con la representación adquirida hasta lograr un canto propio de la especie [Konishi, 1965, Konishi, 1978].

Esta característica de aprendizaje en etapas dependiente de la audición para lograr vocalizaciones maduras propias de la especie presenta muchas similitudes con la adquisición del habla en humanos. Comparte asimismo que en ambos casos el comportamiento aprendido son señales acústicas de alta complejidad con características similares en varias oportunidades [Doupe and Kuhl, 1999]. Es pertinente señalar la diferencia entre un aprendizaje auditivo y un aprendizaje vocal. En el primer caso el animal es capaz de realizar asociaciones con determinados sonidos percibidos y en el segundo, el animal es capaz de generar sonidos adquiridos que no podía realizar de forma innata. Por lo general este aprendizaje es sobre sonidos de su misma especie [Marler, 1967]. Por lo tanto, dada esta definición, la capacidad de aprendizaje vocal es una característica no tan difundida dentro del reino animal. De hecho el único primate que tiene dicha capacidad es el humano, y en el resto de los mamíferos solamente algunos cetáceos y murciélagos. Sin embargo,

en la aves, de las aproximadamente 10,000 especies conocidas, más del 40% de estas aprenden a vocalizar de esta manera, distinguiéndose tres grande grupos el suborden Oscinos (llamadas “aves canoras” como el Canario, Diamante Mandarín, Zorzal, etc), los loros y los colibríes [Jarvis, 2004].

Esta particular característica compartida de aprendizaje sensorimotor para controlar un sofisticado órgano vocal sugiere que los mecanismos neuronales involucrados en dicho proceso sean también compartidos entre humanos y aves. Efectivamente estudios empleando diversas técnicas experimentales mostraron la existencia de estructuras neuronales con funciones análogas en los tres grupos de aves y similares a las conocidas en humanos [Jarvis, 2004].

De esta forma, el canto de aves se consolida como un excelente modelo animal para estudiar circuitos neuronales compatibles con el aprendizaje y la generación de comportamientos complejos. Es por ello que gran parte de las investigaciones en el área se focalizaron inicialmente en conocer cómo es el circuito neuronal del canto: cuáles son los núcleos involucrados, cuáles son sus características y conectividades.

1.2. El sistema del canto

Se conoce en la literatura al sistema del canto como el circuito neuronal específico encargado de la producción y aprendizaje vocal. Estudios comportamentales sugieren que deben existir tres circuitos neuronales distintos encargados de procesos específicos [Brainard and Doupe, 2002]:

- **Vía Motora:** encargada de producir las instrucciones que controlan el órgano vocal capaz de producir la variedad de sonidos del canto.
- **Vía Auditiva:** encargada de percibir y memorizar los sonidos del tutor.
- **Circuito AFP:** encargado de evaluar la realimentación auditiva, compararla con el templado interno y generar las instrucciones de corrección necesarias.

Para comprender cómo funciona se realizaron diversos estudios que fueron desvelando cuál era la neuroanatomía del cerebro de las aves. Es decir, cuáles eran

las estructuras participantes, cómo se encontraban organizadas y cuáles eran las posibles funciones dentro del sistema del canto. Las descripciones neuroanatómicas dieron lugar a interpretaciones neuroetológicas, i.e cómo esas estructuras y esa organización dada son capaces de generar en su conjunto estímulos nerviosos que, controlando un órgano vocal sofisticado finalmente se traducen en el comportamiento complejo observado: el canto.

1.2.1. Descripción neuroanatómica

- **Vía Motora:** El canto al ser un comportamiento complejo requiere un sofisticado control motor que permita coordinar los músculos del órgano fonador con los músculos respiratorios. Es por ello que es necesaria la existencia de una vía encargada de dar las instrucciones adecuadas. En los primeros estudios (en canarios) que se realizaron en esa dirección se lesionaron áreas específicas del cerebro y se evaluó posteriormente cómo influía en la generación del canto. Así se encontró que existían núcleos organizados jerárquicamente que eran indispensables para la generación de canto. En particular dos de los núcleos más relevantes de la vía motora, el HVc (usado como nombre propio) y el RA (*Robustus nucleus of the arcopalium*) al ser lesionados generaban canto anormal o ningún canto [Nottebohm et al., 1976].

Luego se realizaron mediciones con técnicas electrofisiológicas para medir la actividad neuronal y se observaron propiedades más específicas del funcionamiento de cada uno de estos núcleos. Se observó que la actividad del RA presenta una estructura temporal precisa y fuertemente correlacionada con característica acústicas del canto. Es decir que la actividad de las neuronas de este núcleo presentan picos de actividad estereotipados durante el canto [Chi and Margoliash, 2001]. Asimismo, se observó que la actividad de neuronas del HVc durante el canto es regular, precisa y a su vez estructurada en disparos cortos [Yu and Margoliash, 1996]. Por otro lado, también con técnicas de electrofisiología se probó que en el HVc se produce la interacción entre los elementos de la Vía Auditiva con la Vía Motora [McCasland and Konishi, 1981].

Las conexiones de la Vía Motora se esquematizan con trazos negros en la figura 1.1, donde se observa que el HVc se conecta de forma directa con el RA. Éste a su vez inerva al núcleo de motoneuronas nXIIts (parte traqueo-siringea del núcleo hipogloso) y a los núcleos RAm (retroambigualis nucleus) y PAm (parambigualis nucleus) [Spiro et al., 1999]. El primer núcleo controla los músculos del órgano fonador aviar (la siringe); el segundo y el tercero controlan los músculos respiratorios durante el canto [R.A.Suthers et al., 1999]. En la figura 1.1 se observan otros dos núcleos: el NIf (núcleo interfaccialis) que se cree que forma parte de la Vía Motora, sin embargo aún no es claro el rol que cumple; el DM (dorsomedial nucleus of the intercollicular complex) estructura presente en aves no oscinas y oscinas que aparentemente se encarga de controlar las vocalizaciones innatas.

- **Vía Auditiva:** El circuito auditivo tiene su origen en un núcleo conocido como *Field L*, que inerva a todas las áreas involucradas en la audición y se conecta también con el HVc [Kelley and Nottebohm, 1979]. Es análoga a la corteza primaria auditiva en mamíferos. La vía auditiva guarda, junto con otras áreas del sistema del canto, una de las características más sobresalientes del canto de pájaros: la respuesta neuronal selectiva de cada individuo al sonido de su canto propio [Margoliash and Fortune, 1992], [Margoliash, 1983]. Se puede observar representada en verde la Vía Auditiva en la figura 1.1 y como se relaciona con los demás circuitos en el sistema del canto.
- **Circuito AFP:** El circuito AFP (*Anterior forebrain pathway*) es un conjunto de núcleos conectados organizados de forma circular que conectan indirectamente los núcleos motores HVc y RA. La principal función de este circuito fue descrita por Bottjer [Bottjer et al., 1984] quien lesionó alguno de los núcleos y observó que el ave juvenil se encontraba incapacitada para aprender las vocalizaciones adultas. No obstante en aves adultas lesiones en este circuito generan degeneraciones del canto a lo largo del tiempo. Este circuito aparentemente participa en la evaluación de la realimentación auditiva del ave durante la etapa de ajuste de las capacidades motoras [Brainard

and Doupe, 2000].

El AFP está compuesto por tres núcleos interconectados como indican las líneas rojas de la figura 1.1: el Área X, DLM (medial nucleus of the dorsolateral thalamus) y LMAN (lateral magnocellular nucleus). Como se observa también en la figura 1.1 el AFP se conecta con la Vía Motora, el HVc proyecta al Área X, realiza el círculo que interconecta los tres núcleos y finalmente el LMAN proyecta sobre RA. Se realizaron diversos estudios lesionando LMAN o Área X y en ambos casos por pérdidas de funcionalidades distintas el resultado fue el mismo: el juvenil no puede desarrollar su canto adulto.

Anatómica, fisiológicamente y por su función el AFP presenta similitudes con los ganglios de base tálamo-corticales en mamíferos, pero con simplemente tres núcleos [Luo et al., 2001]. La combinación de un circuito específico de ganglios basales con una demanda de sostener un emergente motor altamente estereotipado, hace que el AFP sea un circuito excelente para revelar cuales son los aportes de los ganglios basales en la función del aprendizaje motor [Brainard and Doupe, 2002].

1.2.2. Interpretación neuroetológica

Un paradigma establecido: “El modelo del reloj”

Los conocimientos adquiridos hasta mediados de la década del 2000 condujeron a la aparición de una interpretación neuroetológica que se instaló en el campo como el paradigma de funcionalidad del sistema motor. Esta interpretación es conocida como el “modelo del reloj” y esencialmente es soportada por los trabajos de M. Fee. Este modelo plantea que la Vía Motora del sistema del canto funciona “río abajo”, organizada jerárquicamente donde los núcleos “río arriba” controlan los núcleos subsiguientes. De esta manera el primer núcleo de la cadena, el HVc, funciona de forma pre-motora encargándose de iniciar la cadena de instrucciones neuronales. Más aún, propone que el HVc funciona realizando una secuencia de disparos equidistantes durante el canto (de allí la idea de “reloj”) [Hahnloser et al., 2002].

Como mencioné en la sección anterior estudios con técnicas de electrofisiología

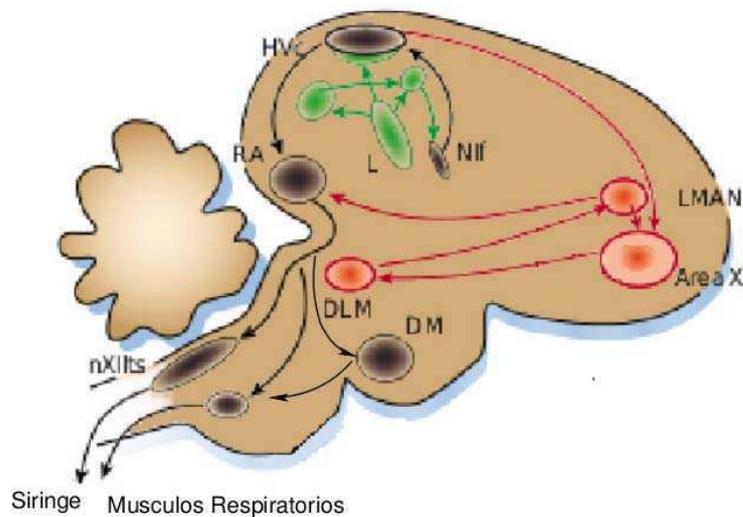


Figura 1.1: Se muestra un esquema del sistema del canto aviar. En código de colores se diferencian la conectividad de las tres vías. En negro se muestra la Vía motora, en rojo el AFP y en verde la Vía auditiva, a su vez se muestran las interconexiones entre cada una de ellas. Figura adaptada de [Brainard and Doupe, 2002]

mostraron que RA presenta una secuencia de disparos muy estereotipada vinculada al canto. Esa observación condujo a que se le asignara un rol pre-motor a este núcleo, asumiendo por lo tanto que se encarga de generar, en parte, la codificación temporal de las instrucciones motoras que en definitiva, se traducen en determinadas características acústicas [Chi and Margoliash, 2001]. Luego, el trabajo de Fee fue un poco más lejos, y en [Hahnloser et al., 2002] muestran mediciones electrofisiológicas en HVC en Diamante mandarín donde observan que a cada disparo en RA le corresponde un conjunto de neuronas de HVC que disparan exactamente a ese tiempo. Por lo que conjeturan que cada conjunto de neuronas activas en un determinado tiempo en RA es controlado por un conjunto de neuronas en HVC. Concluyendo que el HVC como población presenta una representación temporal que cubre todo el tiempo de la canción con disparos equidistantes, que en la analogía con el reloj serían los “tics” [Hahnloser et al., 2002, Fee et al., 2004].

Me interesa resaltar dos de los conceptos más relevantes que instala este paradigma:

- **Rol pre-motor del HVC y RA:** el rol pre-motor que le asigna a los núcleos HVC y RA suponiendo un sistema que funciona “río abajo”. Esto implica que las instrucciones neuronales que devienen en características acústicas se generan en los “primeros” núcleos (HVC con “tics“ y RA controlado por HVC generando disparos vinculados a la canción). Eso implica que los demás núcleos de la cadena motora simplemente replican la instrucción hasta llegar a los músculos.
- **HVC como un reloj:** disparos cortos que generan un continuo de “tics” que cubren regularmente toda la duración del canto, independientemente de las características del mismo.

Remarco estos conceptos que impone el “modelo del reloj” pues fueron ampliamente aceptados y establecidos en la comunidad de canto de aves durante muchos años. Sin embargo, en los últimos años comenzaron a aparecer nuevos trabajos que plantean la posibilidad de otros paradigmas posibles para interpretar el sistema del canto en aves.

Nuevos indicios asoman

Un trabajo relevante que observa un comportamiento de disparos diferente al modelo del reloj en HVC es el trabajo de R. Mooney. Éste encontró que en gorriónes negros la respuesta de la neuronas de HVC proyectoras al área X (HVC_x) sucede de forma precisa y responde a la presentación auditiva de una secuencia de notas determinada [Prather et al., 2008]. Otro trabajo relevante en esa dirección es el trabajo que compone el cuerpo central de esta tesis doctoral [Amador et al., 2013] (trabajo que será contado en detalle en capítulos subsiguientes). Allí se comprobó a partir de mediciones en Diamante mandarín que los disparos de las neuronas de HVC proyectoras a RA (HVC_{RA}) son sincrónicas con momentos precisos de los denominados gestos motores.

Por el otro lado, surgieron trabajos que ponen en duda la estructura “río abajo” del sistema del canto en aves. Uno de ellos es el de R. Ashmore y Schmidt [Ashmore, 2008] quienes mediante técnicas de electrofisiología estimularon diversos núcleos, ubicados “río abajo” y observaron cuál era el recorrido de dicho estímulo

del canto en aves suboscinas, aves oscinas y palomas. La figura 1.3 muestra como los núcleos DM, RAm, PAm y nXII se encuentran presentes en los tres casos, y en todos ellos las aves son capaces de fonar. Por lo tanto este resultado sugiere que la vía motora se iniciaría por debajo de los núcleos RA y HVC, quedando éstos vinculados al aprendizaje con conexiones “río arriba” desde la parte motora que sería modulada asimismo por estos núcleos [Liu et al., 2013]. Este resultado también resulta consistente con uno de los experimentos planteado por D. Vicario en [Vicario and Simpson, 1995] donde estimula eléctricamente DM y obtiene vocalizaciones sencillas, mostrando que ese conjunto de núcleos es suficiente para fonar, pero no para aprender.



Figura 1.3: Un esquema aproximado que representa la estructura neuronal de control motor en tres ordenes distintos de aves con diferentes capacidades de aprendizaje. Se observa como las aves que aprenden como los oscinos presentan las estructuras superiores del sistema del canto vinculadas al aprendizaje (AFP, Vía auditiva y parte superior de la Vía motora) y la parte inferior del sistema vinculada generación del canto. Las aves suboscinas que no aprenden conservan la parte inferior de la Vía motora, los núcleos que controlan los músculos respiratorios y siringeos, pero no así la parte superior del sistema del canto, donde se observan pseudo-núcleos que podrían relacionarse con algunos núcleos de la parte superior como el RA y el HVC. Asimismo las palomas, que tampoco aprenden a cantar, simplemente conservan los núcleos de la parte inferior del sistema sin mostrar otros núcleos vinculados al canto. Figura adaptada de [Liu et al., 2013]

A partir de estos trabajos se retomó el debate acerca del funcionamiento del sistema del canto y sus principales núcleos. En particular en esta tesis voy a contar el debate que propone la aparición del “modelo de gestos motores” como modelo alternativo al del reloj.

Un debate científico: “Modelo de reloj“ vs. ”Modelo de gestos motores“

Como primer interrogante al momento de comprender cómo codifica el sistema motor, surge la necesidad de comprender entonces cómo se segmenta el canto de aves, cuáles son las unidades funcionales del mismo. El comportamiento evidencia que el canto estaría segmentado en grupos de 100 – 250 *ms* llamadas sílabas. Sin embargo las mediciones electrofisiológicas en HVc sugieren otro tipo de unidad basándose en la representación neuronal que codifica el canto que no necesariamente está vinculada con las sílabas identificadas en el comportamiento. Es allí, en esta representación, donde se plantea el debate. Distintos grupos obtienen mediciones de la actividad neuronal en HVc durante el canto y se preguntan *Cuál es la unidad de canto que codifica?, Cuál es una buena representación del canto para entender la actividad del HVc?*

Por un lado el ”modelo de reloj“ al sostener que el HVc actúa como tal, generando un continuo de disparos en el tiempo, propone que la unidad del canto en la codificación neuronal son los ”tics“ de esta regla temporal sin correlación con ninguna característica acústica del canto. Por lo tanto las mediciones obtenidas en HVc durante el canto para el grupo de M. Fee son algunos ”tics“ del reloj pertenecientes al continuo de ”tics“ [[Hahnloser et al., 2002](#)].

Por el otro el ”modelo de los gestos motores“ propone una unidad del canto distinta, más corta y específica que la sílaba, pero no independiente del canto: los gestos motores. Para ello nuestro trabajo se basó en el desarrollo durante años de modelos matemáticos de baja dimensión de la dinámica del órgano fonador aviar. Esto nos permitió realizar un mapeo de coordenadas acústicas en coordenadas motoras encontrando así una buena representación para el canto. De esta manera, el canto queda segmentado en gestos motores que se traducen en características acústicas que componen las sílabas y luego la totalidad del canto. En esta representación del canto pudimos observar que los disparos de las neuronas del HVc estaban temporalmente sincronizados con los extremos de estos gestos motores, i.e. que las neuronas del HVc estarían codificando el canto en esta representación motora y no funcionando como un reloj autónomo del canto [[Amador et al., 2013](#)].

Estas interpretaciones encontradas están muy bien planteado en el artículo de

Cap.1 Introducción

Todd Troyer donde instala el debate en el área y comenta que puede perdurar por un tiempo, ya que reabrió la pregunta sobre el sistema motor en aves, sistema que se consideraba ya conocido [Troyer, 2013]. Este debate y el "modelo de gestos motores" serán abordados en profundidad en el cuerpo central de esta tesis. La figura 1.4 es un resumen gráfico de lo que postulan ambos modelos, suponiendo alguna actividad medida en HVC.

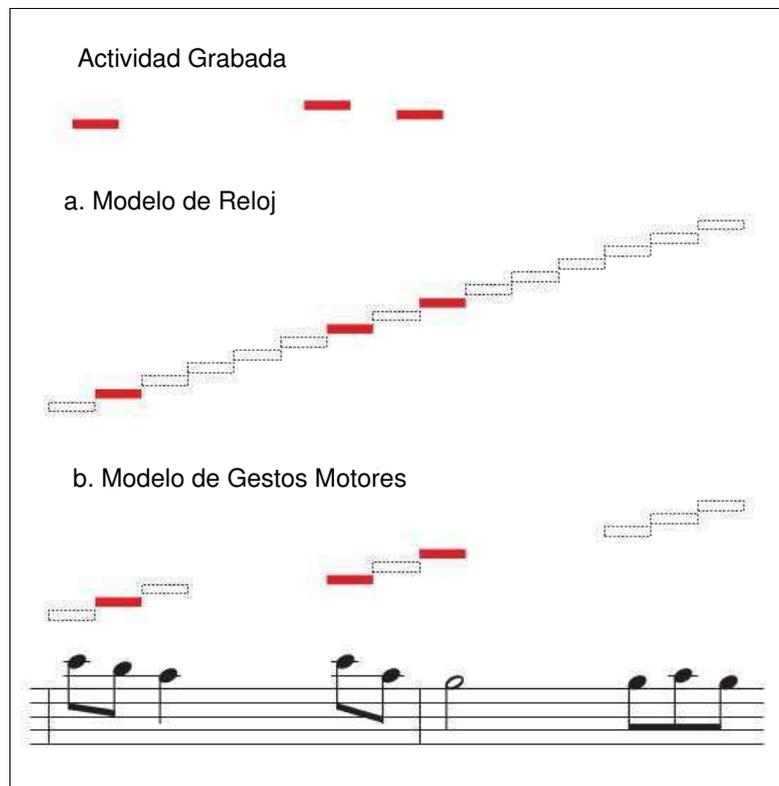


Figura 1.4: Se realiza una ilustración con un pentagrama musical que da cuenta de lo que proponen tanto el modelo del reloj como el modelo de los gestos motores. En la parte superior un esquema de actividad neuronal grabada y en la parte inferior como dicha actividad es interpretada por cada modelo considerando la totalidad de un canto representado por el pentagrama inferior. Figura adaptada de [Troyer, 2013]

1.3. Una biomecánica descriptiva

Los estudios en neurociencia vinculados al sistema del canto, no pueden realizarse sin entender como funciona en su totalidad el canto de aves. Esto implica, no se puede entender qué tipo de instrucciones genera el cerebro, cómo se vinculan con el canto y cómo codifica el sistema nervioso sin antes tener un amplio conocimiento del funcionamiento del órgano vocal efector. En particular, la Siringe, es un órgano complejo altamente no lineal controlada por un conjunto de músculos [Fee, 2002]. Es por ello que numerosas investigaciones se encargaron de comprender biomecánicamente a la Siringe.

Existen dos formas complementarias de abordar el estudio de la Siringe: una biomecánica descriptiva y una biomecánica cuantitativa. La primera se basa en estudios fisiológicos, relevando cuál es la anatomía y como actúan los músculos sobre ésta. La segunda se basa en el modelado matemático de los elementos involucrados en la generación del canto, la Siringe y el tracto vocal superior. En esta sección voy a ocuparme de compartir los conocimientos tenidos hasta el momento desde un abordaje descriptivo, basándome en una revisión de la literatura. En el capítulo siguiente me dedicaré a fondo al estudio cuantitativo, cuyo principal desarrollo es responsabilidad de las investigaciones realizadas en el Laboratorio de Sistemas Dinámicos durante la última década.

1.3.1. Generación de sonido.

El canto de aves esconde un sistema biomecánico capaz de oscilar y generar sonido. Ese mecanismo consiste en un flujo de aire controlado por la presión de los sacos que aéreos atraviesa una válvula que modula el mismo flujo. Dicho mecanismo es similar al que empleamos los humanos para generar el habla [Mindlin and Laje, 2005]. Luego entre la fuente sonora y el entorno aparece el tracto vocal superior, que son los órganos que atraviesa el flujo de aire una vez que sale de la fuente y antes de emerger del pico hacia entorno. Por lo tanto el tracto vocal superior genera un filtrado sobre el sonido que aporta características tímbricas al canto según la anatomía de cada especie [Riede et al., 2006].

Vale aclarar que existen otros mecanismos conocidos en aves para la generación de sonido como puede ser el silbido, propuesto principalmente para la

generación de sonidos tonales [Nottebohm et al., 1976]. Sin embargo, el mecanismo establecido y aceptado para la mayoría de las especies es el de la válvula ubicada en un sector de la vía respiratoria. Esta válvula cuenta con un par de membranas que al ser aducidas hacia el lumen de la vía son puestas en movimientos al ser atravesadas por el flujo de aire. Luego las propiedades mismas de estas membranas, como ser su elasticidad, su tensión, su restitución, generan que el tipo de movimiento al que son inducidas tenga características propias. Por ende, la forma en la que se mueven estas membranas a su vez modula al mismo flujo de aire responsable de ponerlas en movimiento [Mindlin and Laje, 2005]. Finalmente si el movimiento resultante de esta interacción entre las membranas de la válvula y el flujo de aire que la atraviesa es una oscilación con frecuencia audible, se genera sonido y su desarrollo en el tiempo da lugar al canto. Muchos modelos matemáticos describen esta relación entre el flujo de aire y las membranas logrando una interpretación cuantitativa y predictiva las oscilaciones generadas mediante este mecanismo [Amador et al., 2008, Sitt et al., 2008].

Ahora bien, comprender un mecanismo compatible con la generación del sonido parece aún lejano a la comprensión de la generación de un canto complejo. Acortar esa brecha es parte de los esfuerzos que viene realizando el Laboratorio de Sistemas Dinámicos. Para ello se encarga de investigar una de las especies más utilizadas en el canto de aves, el Diamante mandarín y en ella se basan los estudios realizados en esta tesis. Por lo tanto para lograr dicho objetivo resulta imprescindible conocer el canto del Diamante mandarín, la anatomía de la fuente sonora, los mecanismos que controlan dicha fuente para generar variaciones en el sonido y como afecta el tracto vocal superior al sonido resultante.

1.3.2. El canto del Diamante Mandarín

El Diamante mandarín es una de las aves canoras más investigadas ya que es una especie que se encuentra muy bien domesticada. Es un ave de tamaño pequeño que tiene la capacidad de reproducirse en todas las épocas del año y se adaptada sin problemas mayores a convivir en colonias con otros animales de su especie en cautiverio.

El macho del Diamante mandarín es capaz de realizar dos tipos de vocaliza-

ciones: el canto de la especie y llamadas ("calls") de advertencia territorial [Thorpe, 1961]). El comportamiento de estas aves indica que los machos realizan el canto como forma de cotejo hacia la hembra (*canto dirigido*) o bien con alguna diferencia sutil a modo de práctica (*canto no dirigido*) [Woolley and Doupe, 2008]. Por ende, al reproducirse todo el año cantan durante todo el año y por ello son un gran modelo animal para estudiar canto de aves.

Como ya he contado, cada individuo de esta especie aprende a cantar mediante un tutor en dos etapas cuando juvenil y aprende un canto. Una vez aprendido este individuo cristaliza un canto propio estereotipado, funcionando como una "huella digital" de este ave. Es decir, las sílabas y por ende las características acústicas del canto son las propias de la especie, pero la articulación de las mismas para formar el canto es propia de cada individuo.

Las características acústicas que presenta el canto del Diamante mandarín son muy variadas a lo largo del canto. La figura 1.5 muestra el sonograma de un canto típico. El eje horizontal representa el tiempo, el eje vertical muestra las frecuencias y en tonos de grises la intensidad de cada valor de frecuencia. Por lo tanto este tipo de representación tridimensional para el canto resulta muy útil ya que permite analizar en el tiempo como evolucionan las distintas componentes armónicas del canto.

Se puede observar que el canto presenta sílabas bien diversas como algunas con alto contenido espectral y baja frecuencia fundamental (cerca de 500 Hz) y otras con bajo contenido espectral y alta frecuencia fundamental (de 4000 Hz hacia arriba). También presenta sílabas ruidosas, donde casi la totalidad de espectro se encuentra con energía. La figura 1.6 muestra esta relación entre frecuencia fundamental y contenido espectral en la señal temporal de la presión sonora y en la transformada de Fourier de la misma.

Esta particular relación aparece como una restricción biomecánica que presenta el órgano fonador. Las formas en las que oscilan los labios limitan a ciertos sonidos caracterizados acústicamente por esa relación entre frecuencia fundamental y contenido espectral [Sitt et al., 2008, Sitt et al., 2010]. Esta observación da lugar a la pregunta sobre cuán complejas deben ser las instrucciones del sistema nervioso para controlar el canto, cuando gran parte del mismo está determinado por la biomecánica. Sobre esta pregunta se focaliza gran parte del trabajo de esta

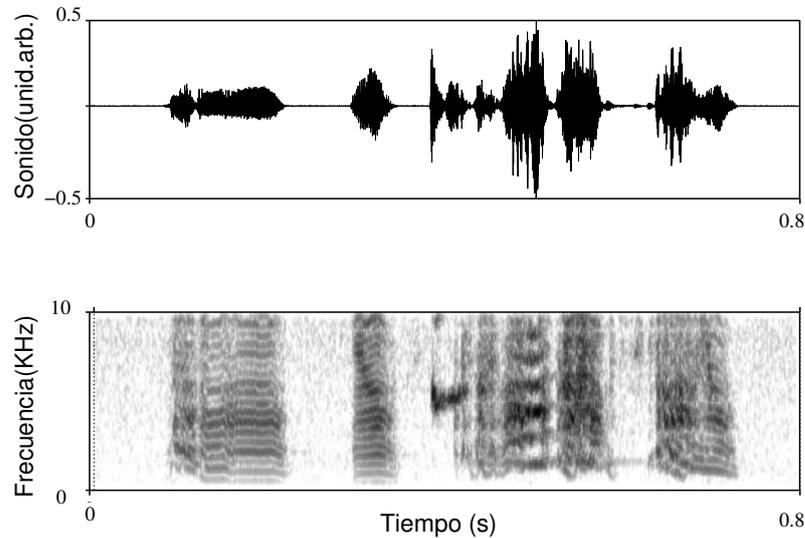


Figura 1.5: Se muestra en el panel superior la señal temporal de la onda de presión del sonido del canto y en el inferior un sonograma típico de un canto de Diamante mandarín. El segundo es una forma ampliamente empleada a la hora de mostrar señales sonoras ya que muestra valores de frecuencias (eje y) con sus correspondientes amplitudes (escala de grises) y cómo evolucionan éstas en función del tiempo (eje x). Se puede notar la presencia de varios tipos de sílabas y como se suceden para formar una secuencia que llamamos motivo del canto.

tesis.

1.3.3. Morfología de la Siringe.

La siringe ha sido un foco de estudio importante en el canto de aves. Inicialmente las distintas morfologías de este órgano fueron utilizadas para clasificar taxonómicamente a las aves [Suthers, 2001, Elemans et al., 2003]. La principal caracterización morfológica que distingue a las clases de siringe, es su ubicación en la vía respiratoria, dando por resultado la siguiente clasificación:

- **Siringe traqueo-bronquial:** las estructuras vibrantes, labios siringeos, se ubican en los anillos cartilaginosos ubicados en la juntura entre la traquea y los bronquios. Esto implica la existencia de dos fuentes sonoras, una en cada juntura traqueo-bronquial, siendo esta clase de siringe una estructura

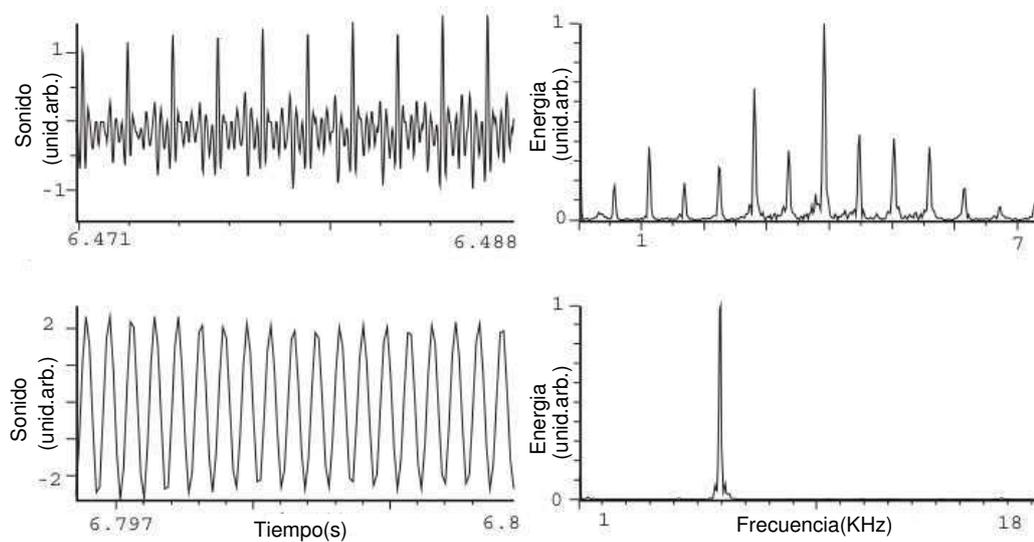
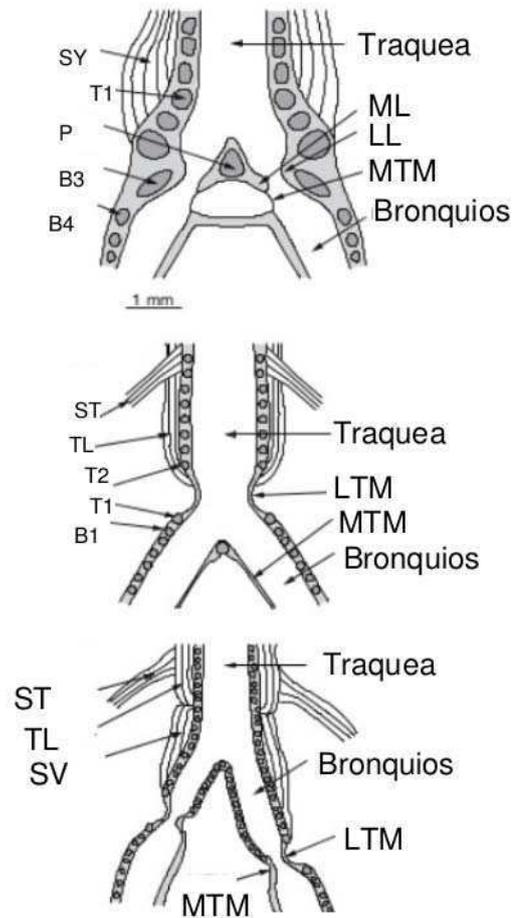


Figura 1.6: Se muestra en el panel superior la señal temporal típica de un sonido de baja frecuencia con alto contenido espectral, que representada en el dominio de las frecuencias se ve como una sucesión de picos equiespaciados con distinta amplitud como se muestra a la derecha. En el panel inferior se observa la señal temporal de un sonido de alta frecuencia fundamental y bajo contenido armónico, asimismo en a su derecha se muestra este tipo de sonido en el dominio de la frecuencia representado como un único pico. Figura adaptada de [Sitt et al., 2008]

bipartita. La mayoría de las especies del suborden Oscinos, las aves canoras, presentan este tipo de estructura siringea. En particular el Diamante mandarín que en párrafos siguientes será descrita en detalle. En la figura 1.7a) se puede observar un sección ventral.

- **Siringe traqueal:** las membranas oscilantes se encuentran únicamente ligadas a anillos cartilagosos traqueales, dando por ende una única fuente sonora. Es muy común en palomas y loros. En la figura 1.7b) muestra una sección ventral.
- **Siringe bronquial:** los labios siringeos involucran anillos cartilagosos bronquiales y es común en Caprimulgiformes, Cuculiformes y Strigidae (Búhos) [Casey and Gaunt, 1985]. Se muestra en la figura 1.7 c).

Figura 1.7: Se muestran las secciones ventrales de los 3 tipos de siringes descritas. a) Siringe traqueo-bronquial, con dos fuentes sonoras muy común en oscinos. Cada una de estas fuentes sonoras está compuesta por un par de labios, labio lateral (LL) y labio medio (ML) y se ubican en la juntura entre los bronquios y la traquea. Las letras T y B representan los anillos traqueales y bronquiales respectivamente; SY, es un músculo siringeo y P, el ossified pessulus; MTM es la membrana media timpanoforme. b) Siringe traqueal, la constricción del lumen traqueal debido a la membrana lateral timpanoforme (LTM) genera la única fuente sonora y es común en loros y palomas. Las letras TL y ST representan los músculos siringeos. c) Siringe bronquial, las válvulas se ubican en los bronquios y presentan distintas distancias entre estas y la traquea generando una siringe asimétrica, es común en Caprimulgiformes, Cuculiformes y Strigidae (Búhos). Figura adaptada de [Mindlin and Laje, 2005]



La principal diferencia que presenta la **siringe en oscinos** respecto al órgano fonador en humanos es su estructura bipartita. Es decir, la laringe en humanos es una única estructura vibrante, mientras que en las aves oscinas, como he dicho, aparecen dos juegos de labios oscilantes uno por cada juntura traqueo-bronquial. Sin embargo cada estructura en sí presenta mucha similitud con la laringe, ya que se trata de membranas de tejido conectivo capaz de oscilar bajo ciertas condiciones adecuadas [Goller and Larsen, 1997].

Las membranas involucradas en la siringe son dos, *Medial Labia* (ML) y *Lateral Labia* (LL) se muestran en el corte ventral de la figura 1.7 a). Estas se encuentran a ambos lados del lumen bronquial y funcionan como una válvula. Mediante la acción de músculos aductores esta válvula es capaz de cerrarse y posicionar a los labios cerca de forma que al atravesar el flujo de aire estos comienzan a oscilar. También puede cerrarse completamente impidiendo el flujo de aire a través de la misma [Mindlin and Laje, 2005]. La parte medial de esta válvula se divide en dos tejidos, por un lado el *Medial Labium* y por otro un tejido fino *Medial Tympaniform Membrane* (MTM) ubicado debajo en el espacio inter-bronquial [Fee, 2002]. Vale remarcar que durante muchos años se pensaba que el tejido responsable de general sonido era el MTM, sin embargo los trabajos de F. Goller a finales de los '90 en palomas y Diamante mandarín, demostraron que no participaba en la fonación. Luego identificó que efectivamente las membranas oscilantes eran los labios LM y ML.

La identificación del tejido responsable del sonido resulta un paso muy importante al momento de comenzar a comprender el sistema. Es decir, recién después de identificar que parte del sistema es la fuente sonora propiamente dicha se puede dilucidar cuáles son los mecanismos que utiliza para la generación de sonido. Más aún, se puede ir más allá y preguntarse cómo es que esa fuente sonora es capaz, no solo de oscilar generando sonido, sino de generar un comportamiento que se modifica en el tiempo presentando precisión en su duración y contenido espectral. En esa dirección gran parte de los esfuerzos se focalizaron en identificar qué tipo de estructura era capaz de alterar esta fuente de forma consistente con el comportamiento observado. Los trabajos de R. Suthers y F. Goller develaron el rol de músculos siringeos durante la generación del canto, siendo éstos responsables de alterar distintas las características de la anatomía de los labios.

1.3.4. Los músculos involucrados

En dos trabajos muy relevantes F. Goller y R. Suthers develaron cuál era el rol de los principales músculos siringeos. Por un lado estudiaron los músculos que aducen y controlan la apertura-cierre de los labios [Goller and Suthers, 1996b] y por el otro los que controlan la siringe durante el canto [Goller and Suthers, 1996a].

Estos autores encontraron que los labios tenían una posición de apertura en la cual dejan pasar todo el flujo de aire y otra de cierre en el cual obturan la vía respiratoria por ese canal. Entre esos dos extremos aparece una tercera posición intermedia que es de fonación, donde el flujo de aire al pasar por la válvula puede provocar oscilaciones en los labios. En [Goller and Suthers, 1996b] observaron que los músculos encargados de la separación de los labios son el ventral traqueo-bronquial ventralis (vTB) y el traqueolateralis (TL). Mientras que los encargados de aproximar los labios son los músculos siringealis dorsalis (dS) y el dorsal traquobronquial (dTb). Para identificar estas funciones observaron la actividad muscular de los primeros dos crecía en la inspiración y la de los segundos en la expiración.

Por otro lado encontraron que la actividad del músculo siringealis ventralis (vS) presentaba una fuerte correlación con la frecuencia fundamental del canto. Este músculo es el encargado de controlar la tensión de los labios, por lo tanto se identificó como el encargado de controlar a la siringe durante el canto [Goller and Suthers, 1996a]. La existencia de un músculo que correlaciona con la tensión labial dio lugar a la presencia en los siguientes modelados de la siringe de un parámetro que de cuenta de ello y sea capaz de controlar la frecuencia del canto [Sitt et al., 2008, Amador et al., 2008, Sitt et al., 2010]. En la figura 1.8 se observa la anatomía muscular de la siringe que presenta un grupo de seis músculos. Vale aclarar que en la especie estudiada al presentar una estructura siringea bipartita, cada una de ellas cuenta con un conjunto de seis músculos independientes.

Los músculos que controlan el funcionamiento de la siringe se pueden separar en dos grandes grupos, los intrínsecos y los extrínsecos. Los primeros tienen su origen e inserción dentro de la siringe controlando la posición relativas de las distintas componentes de la misma. Mientras que los segundos conectan la siringe

con elementos exteriores, y controla a la misma como un todo [Larsen and Goller, 2002]. Estudios morfológicos de la siringe dan cuenta que los músculos intrínsecos al permitir el control independiente de distintas componentes de la siringe, son propios de las aves canoras. No así los músculos extrínsecos que se observan en todos los tipos de aves.

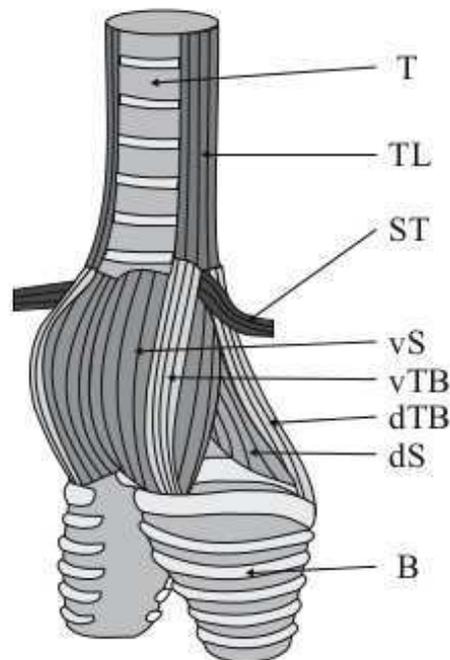


Figura 1.8: Se muestra una vista ventro-lateral de la siringe indicando los músculos más relevantes en la generación del canto. El músculo siringealis ventralis (vS) controla en gran parte la frecuencia fundamental de la fonación; los músculos siringealis dorsalis (dS) y el dorsal traqueo-bronquial (dTB) se encargan del cierre de los labios; los músculos ventral traqueo-bronquial (vTB) y el traqueolateralis (TL) se encargan de la separación entre los labios; el otro músculo es extrínseco (conecta la siringe con órganos fuera de ésta) es el sternotraquialis (ST); las letras T y B representan la tráquea y los bronquios respectivamente. Figura adaptada de [Goller and Suthers, 1996b]

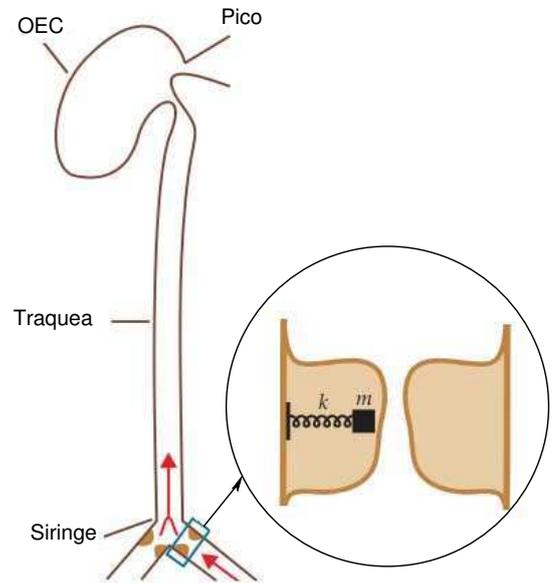


Figura 1.9: La figura muestra un esquema del tracto vocal superior. En la base se muestra la estructura bipartita siringea, conectada con la tráquea que mediante una glotis se comunica con la OEC y el pico.

1.3.5. El tracto vocal superior

Entre la siringe y el entorno se encuentra el tracto vocal superior, que como ya he dicho afecta al sonido resultante. Anatómicamente el tracto vocal superior del Diamante mandarín está compuesto por la tráquea, que se conecta mediante una constricción (glotis) con la cavidad orofaríngea-esofágica (OEC) que luego se comunica vía el pico del ave con el exterior. El sonido que se genera en la siringe atraviesa todos esos componentes del tracto vocal y cada uno de ellos es capaz de modificar sus características anatómicas como ser su largo, ancho, volumen, apertura y cierre modificando así sus efectos sobre el sonido [Fletcher et al., 2006a]. La figura 1.9 muestra un esquema de como se conectan estos elementos entre sí y con la siringe.

Existen dos teorías sobre como interactúa el tracto vocal superior con la siringe. Por una lado la teoría de fuente-tracto desacoplados, que propone que el tracto actúa como un filtro sonoro sobre las oscilaciones resultantes en la siringe debido a su propia dinámica, sin interferir en la misma. Por el otro la teoría de fuente-tracto acopladas, en donde las características del filtro afectan a la dinámi-

ca propia de la siringe dando como resultante otro tipo de oscilaciones.

En la primera interpretación, la siringe mantiene su dinámica independiente de lo que suceda con el filtro y ésta es solo controlada por los músculos y el flujo de aire que la atraviesa. Las diferencias de presión generadas en la siringe se transfieren a la traquea y al resto del tracto que le imprime un filtrado sonoro [Fletcher et al., 2006a]. En algunos casos se estudió que el tracto es controlado de forma tal que resulte un filtro que enfatiza la frecuencia fundamental del canto [Riede et al., 2006, Daley and Goller, 2004]. En otros trabajos se estudio el rol del pico como una componente del filtrado sonoro [Goller et al., 2004]. Asimismo se realizaron estudios donde se hizo cantar al ave en una atmósfera con helio y se observaron los cambios en las características acústicas dando cuenta de la influencia del tracto sobre el mismo [Nowicki, 1987].

La segunda interpretación propone que el acople acústico y mecánico existente entre la siringe y el filtro determina una dinámica distinta. En esa dirección se realizaron diversos modelos matemáticos que dan cuenta de propiedades acústicas que son indicios de este acople [Arneodo and Mindlin, 2009, Arneodo et al., 2011]. Esta interpretación también permite explicar los resultados de los experimentos con helio entre otros, por lo que ambas interpretaciones son válidas según la escala en que sea necesario describir el sistema. En particular durante el desarrollo de esta tesis voy a considerar al tracto como un filtro con dinámica propia que será modelado de manera independiente de la siringe.

Capítulo 2

Una biomecánica cuantitativa

El abordaje complementario al estudio anatómico y funcional del órgano fonador aviar es el estudio, por medio de modelos matemáticos, de la dinámica resultante de la interacción entre los actores descriptos. Esto implica modelar a las membranas oscilantes de la siringe como masas de tejido elástico sometidas a distintas fuerzas: la fuerza de restitución, fuerzas disipativas y la presión que reciben del flujo de aire que las atraviesa. De esta forma en el Laboratorio de Sistemas Dinámicos se vienen desarrollando distintos modelos matemáticos de carácter predictivo a partir del análisis de las fuerzas aplicadas sobre los labios.

Mediante estos modelos se busca reconstruir las características acústicas dando cuenta de las restricciones biomecánicas que impone el sistema. Los modelos predictivos permiten comprender cómo serán las oscilaciones resultantes dependiendo de las condiciones dinámicas en las cuales se inicia el movimiento. Asimismo el modelado es de baja dimensión procurando establecer una jerarquía en las variables del sistema incluyendo solamente aquellas variables relevantes para la generación de las características acústicas.

Durante más de una década se han ido desarrollando diversos trabajos donde implementando modelos matemáticos basados en la fisiología del ave se logró sintetizar canto y vincular las variables del modelo con elementos de la fisiología animal. Los trabajos seminales en esa dirección fueron desarrollados por el Laboratorio de Sistemas Dinámicos, donde en el año 2001 se desarrolló un modelo minimal de la siringe que predijo gran parte de las características del canto

para Canarios (*Senirus canaria*) [Gardner et al., 2001]. En trabajos siguientes se probó la relación existente entre los parámetros que controlan al modelo con las variables biológicas involucradas. En primera instancia se realizaron simulaciones con un modelo para el canto del Chingolo (*Zenotrichia capensis*) y se las interpretó en términos de la actividad de los músculos involucrados [Laje et al., 2002] logrando una gran variedad de sonoridades distintas. Luego se controló el modelo con medidas fisiológicas de los músculos y los sacos aéreos para Cardenales del norte (*Cardinalis cardinalis*) [Mindlin et al., 2003]. Los modelos pertinentes al desarrollo de esta tesis, son los que se desarrollaron para el Diamante mandarín y fueron relatados en trabajos como el de Ana Amador [Amador et al., 2008], el de J. Sitt [Sitt et al., 2008] y del autor de esta tesis [Perl et al., 2011]. En el primero se desarrolló el modelo físico estudiando la dinámica propia del órgano fonador considerando directamente las fuerzas actuantes sobre el mismo; en el segundo se realizó un procedimiento típico en el estudio de sistemas dinámicos y se estudió la forma normal del modelo físico; en el tercero se ajustaron adecuadamente los parámetros de la forma normal para obtener las características acústicas del canto del Diamante mandarín correlacionándolos con señales fisiológicas conocidas.

2.1. El modelo físico

2.1.1. El modelo

Las características acústicas del canto del Diamante mandarín relatadas en el capítulo anterior abrieron la pregunta acerca de cómo una misma fuente es capaz de generar un espectro tan amplio de sonidos. Es allí donde la dinámica no lineal resulta una herramienta adecuada para describir dicha situación. Preguntas del estilo: cómo un par de masas inicialmente en reposo al ser sometidas a una fuerza externa son puesta en movimiento; cómo el movimiento de dichas masa puede ser oscilatorio; qué tipo de oscilaciones pueden realizar, son preguntas que la dinámica puede responder.

Uno de los primeros modelos que se propuso describir cómo son puestas en movimiento membranas al ser atravesadas por un flujo de aire, para el caso de

humanos, fue el de Titze [Titze, 1988]. En este trabajo el autor observó que las cuerdas vocales eran capaces de efectuar dos tipos de oscilaciones: una oscilación lateral (perpendicular al sentido del flujo de aire) y otra vertical a través de la superficie de los labios (paralela al flujo de aire). Asimismo observa que ambos tipos de movimiento no son independientes, sino que actúan de manera coordinada generando las condiciones necesarias para que el sistema gane energía en cada ciclo permitiendo que se produzca un movimiento oscilatorio. Este modelo se lo conoce como el “modelo de aleteo” por su similitud con el movimiento de un ala en un vuelo de ave [Titze, 1988]. Este modelo sencillo logra predecir satisfactoriamente en término de parámetros fisiológicos realistas las frecuencias fundamentales de los sonidos voceados en humanos. Para el caso de aves se observaron experimentalmente movimientos similares lo que permitió la implementación del “modelo de aleteo” en aves [Larsen and Goller, 1999] y se observó la presencia de dinámicas no lineales en la siringe [Fee et al., 1998]. Representa por lo tanto la base para el desarrollo de los modelos matemáticos de canto de aves.

El “modelo de aleteo” propone describir el movimiento de los labios descomponiéndolo en ambos modos. El modo lateral que describe el movimiento del centro de cada masa midiendo el apartamiento de la posición de equilibrio como un todo. El modo longitudinal que es el responsable de establecer una diferencia de fase entre la parte superior e inferior de los labios. Asimismo, como he contactado, establece una relación entre ambos modos de forma de generar oscilaciones autosostenidas. Para ello debe ganar energía en cada ciclo debido al flujo de aire que la atraviesa de forma de compensar las pérdidas disipativas del sistema. Para ganar dicha energía por ciclo, la forma natural de vincular ambos modos es que al separarse la posición media de los labios el perfil sea convergente y al acercarse sea divergente. De esta forma la presión interlabial es mayor cuando los labios se están separando que cuando se están acercando dando como resultado fuerzas no uniformes actuando sobre las membranas que dan una ganancia neta de energía por ciclo. Vale aclarar que para no generar cambios tan dramáticos en el perfil se logra la misma relación de fuerzas no uniformes considerando un perfil menos divergente cuando abre que cuando cierra.

Entonces si se llama a_1 y a_2 a la mitad de la separación entre los labios en la parte inferior y superior respectivamente y x al apartamiento del equilibrio del

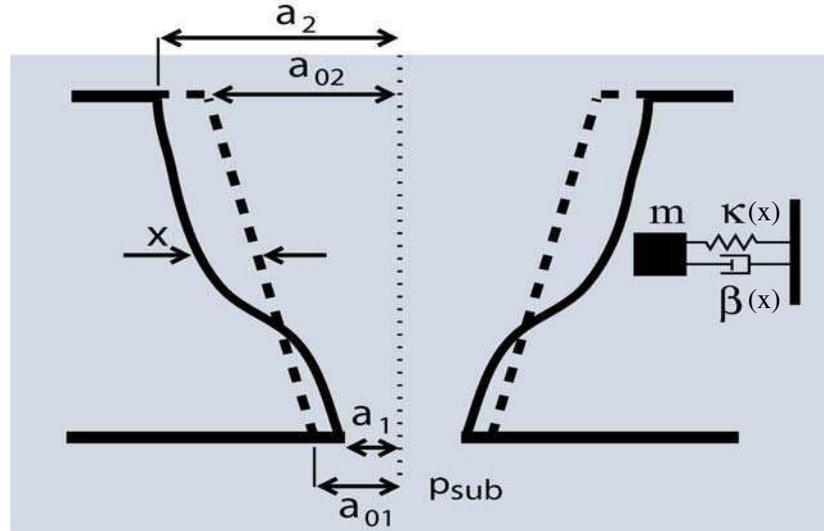


Figura 2.1: Sección frontal de la siringa. Se observa la posición del punto medio de los labios (x) y el significado de las variación que participan en las ecuaciones que rigen su dinámica descrita por la ecuación (2.5).

punto medio de los labios podemos empezar a describir las ecuaciones que rigen el modelo (ver figura 2.1). Ahora si llamamos τ al tiempo que demora la onda longitudinal en recorrer medio longitud del labio, podemos escribir para a_1 y a_2 :

$$a_1 = a_{01} + x + \tau y \quad (2.1)$$

$$a_2 = a_{02} + x - \tau y \quad (2.2)$$

Donde a_{01} y a_{02} son la separación media en el borde inferior y superior respectivamente en el equilibrio. La variable y representa la velocidad del punto medio del labio. Se puede calcular la presión interlabial promedio como [Titze, 1988]:

$$p_{av} = p_{sub} \left(1 - \frac{a_2}{a_1}\right) \quad (2.3)$$

en donde p_{sub} es la presión de los sacos aéreos antes de atravesar la siringa. Por lo tanto con dicha expresión para la presión promedio interlabial se puede

calcular la fuerza f_{lab} ejercida sobre los labios debido a la presión como:

$$f_{lab} = a_{lab}p_{lab} \quad (2.4)$$

Describiendo entonces la dinámica de la posición media de los labios (x) como un todo, considerando el aleteo en la expresión de la fuerza interlabial debido al flujo de aire, se puede escribir un sistema dinámico partiendo de las leyes de Newton y el acción de cada una de las fuerzas sobre el cuerpo de masa m [Amador et al., 2008]:

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \frac{1}{m}(-k(x)x - \beta(y)y - cx^2y + f_0 + a_{lab}p_{sub}(\frac{\Delta a + 2\tau y}{a_{01} + x + \tau y})) \end{aligned} \quad (2.5)$$

El primer término de la segunda ecuación corresponde a una fuerza de restitución no lineal donde $k(x) = k_1 + k_2x^2$. El segundo término de esa misma ecuación es una fuerza disipativa no lineal donde $\beta(y) = \beta_1 + \beta_2y^2$. El tercer término es una fuerza disipativa no lineal que da cuenta de los límites del movimiento de los labios al estar inmersos en una anatomía determinada. Es decir, cuando los labios se acercan mucho, chocan entre sí, cuando se separan demasiado chocan cada uno contra las paredes del lúmen bronqueo-traqueal. El cuarto término, f_0 , corresponde a una fuerza independiente de la dinámica de los labios y modela el cierre activo de los labios [Gardner et al., 2001, Laje et al., 2002]. Finalmente el último término representa la fuerza ejercida sobre los labios por la presión promedio interlabial, donde $\Delta a = a_{01} - a_{02}$.

El resultado de integrar este modelo es una serie temporal $x(t)$ que da cuenta del movimiento de los labios en función del tiempo, el cual en determinadas condiciones, que fueron estudiadas en el trabajo de Amador, es oscilatorio encargándose de modular el flujo de aire. Esta modulación de aire resulta en sonidos que presentan las características acústicas del canto del Diamante mandarín. Para analizar cómo se comporta el modelo ante distintos valores de los parámetros se empleó un recurso típico del estudio de los sistemas dinámicos que es la construcción de diagramas de bifurcaciones identificando cuales eran las variables relevantes en sus efectos dinámicos y su identificación con la fisiología del ave.

2.1.2. Estudio dinámico

En principio se buscó identificar a los parámetros de la ecuación (2.5) con las señales fisiológicas conocidas que controlan la siringe. En esa dirección, el conocer que la contracción del músculo vS produce la extensión de los labios, permitió asumir que la parte lineal del coeficiente de restitución de la ecuación (2.5) (k_1) es proporcional a la tensión de dicho músculo. Asimismo la presión subsiringea (p_{sub}) que controla el flujo de aire que atraviesa la válvula y es responsable de entregar energía necesaria que compense la disipación se presentó como el otro parámetro relevante. Esta identificación fue verificada en distintas especies al generar canto sintético con los modelos reemplazando los parámetros k_1 y p_{sub} por medidas fisiológicas adquiridas en las aves [Mindlin et al., 2003, Sitt et al., 2010].

En el trabajo de Amador se realizó el estudio sistemático de cómo se comporta el modelo ante la variación de los parámetros identificados como relevante p_{sub} y k_1 . Para ello se construyó un diagrama de bifurcación que consiste en identificar y caracterizar a las soluciones del sistema según se encuentren en alguna región del espacio de parámetros (p_{sub}, k_1) determinada. Para ello se realizaron simulaciones del modelo para una grilla de valores de los parámetros y se estudiaron las características de las soluciones y las bifurcaciones del sistema. Las bifurcaciones son aquellos valores de los parámetros donde el sistema cambia cualitativamente sus soluciones [Strogatz, 2001]. Entender de qué manera el sistema cambia de soluciones resulta muy relevante, ya que representa en este caso la forma en que los labios comienzan a oscilar, definiendo entonces cuáles son las características acústicas resultantes de esa oscilación. La figura 2.2 muestra el diagrama de bifurcaciones de este modelo indicando con distintos tipo de líneas las bifurcaciones.

En la región 1 de la figura 2.2 existe un único punto fijo atractor, implica que cualquier condición inicial de la posición de los labios, termina en un punto del espacio de fases (espacio determinado por la posición y la velocidad de los labios (x, \dot{x})). Esta no es por consiguiente una solución oscilatoria que pueda sintetizar canto. En la región 2 aparece una solución oscilatoria sin dejar de existir un único punto fijo. Esto implica que cualquier condición inicial para la posición y velocidad de los labios, termina teniendo un movimiento oscilatorio, como lo

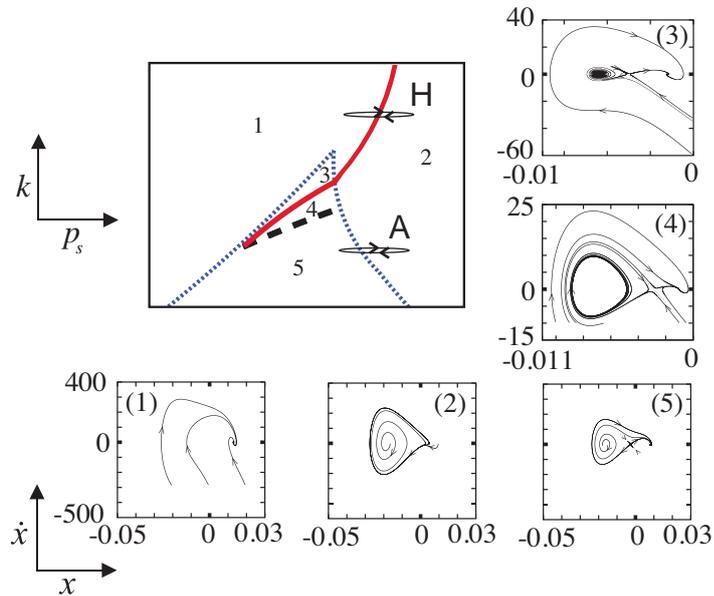


Figura 2.2: Se muestra el diagrama de bifurcaciones y las trayectorias en el espacio de fases según los valores de los parámetros. Éstos son la presión sub-siringea p_{sub} y la parte lineal de la restitución, el coeficiente k , que se asume proporcional a la actividad de los músculos siringeos (vS). Si los valores de los parámetros están en la región 1, sólo existe un punto fijo atractor, por lo que la dinámica tiende a ese punto. En la región 2, aparece un ciclo límite atractor dando soluciones oscilatorias. En las regiones 3,4 y 5 existen 3 puntos fijos coexistiendo con un ciclo límite en la región 4. La línea negra puntuada representa una bifurcación homoclínica, las líneas azules puntuadas representan las bifurcaciones de nodo-silla y la línea roja la bifurcación de Hopf. Las regiones en el espacio de fases y los espacios de fases presentados llevan los mismos números indicando su correspondencia.

muestra el diagrama de fases correspondiente en la figura 2.2. Este movimiento oscilatorio de los labios es capaz de modular el flujo de aire generando finalmente sonido. Para que un punto fijo atractor se transforme en una solución atractora oscilatoria el sistema debió atravesar una bifurcación, indicada con la letra H en la figura 2.2, llamada bifurcación de Hopf. Este bifurcación consiste en que un punto fijo atractor pierda estabilidad generando un repulsor y un ciclo límite atractor. Por lo tanto el movimiento generado vía este tipo de bifurcación son oscilaciones que nacen con frecuencia definida y amplitud cero [Strogatz, 2001]. Esto determina que el sonido generado por el movimiento de los labios sea prácticamente un tono puro, es decir una frecuencia fundamental bien determinada sin

contenido armónico.

En las regiones 3, 4 y 5 coexisten tres puntos fijos, en la figura 2.2 se muestra a las tres regiones limitadas por líneas puntuadas azules que representan una bifurcación de nodo-silla. Esta bifurcación consiste en la aparición, en un primer momento, de un punto fijo (exactamente sobre las líneas) que luego se separa en dos puntos fijos (o viceversa, dos puntos fijos que colapsan en uno y luego desaparece). Al realizar la transición de la región 1 a la región 3, el punto fijo que existía se conserva y nacen dos nuevos puntos fijos. El atractor generado en esta bifurcación al atravesar hacia la región 4 se ve afectado por una bifurcación de Hopf, por lo que aparece un ciclo límite en 4. La separación entre la región 4 y la región 5 está dada por una bifurcación homoclínica (línea negra punteada en la figura 2.2), donde el ciclo límite colapsa con el punto fijo silla. La región 5, presenta una rica estructura, la rama inestable del punto silla es parte de la rama estable del atractor, generando una bifurcación distinta al realizar la transición de 5 a 2 (indicado con la letra A en la figura 2.2): nodo-silla en ciclo límite (SNILC). En este caso los dos puntos fijos colapsan (atraviesa una bifurcación de nodo-silla), pero como la rama inestable de uno coincide con la rama estable del otro, se producen oscilaciones en el sistema. Pero éstas son sustancialmente diferentes a las generadas mediante una bifurcación de Hopf, éstas comienzan con periodo infinito (frecuencia 0) y amplitud finita. Esto se debe a que al atravesar la bifurcación, queda el “fantasma“ del nodo-silla existente, por lo que el labio no oscila uniformemente, sino que se detiene mucho tiempo en las cercanías de donde existía el nodo-silla y luego hace una rápida excursión por el resto de la oscilación. Es por ello que estas oscilaciones, en valores de los parámetros cercanos a la bifurcación de SNILC, generan sonidos con alto contenido espectral [[Amador et al., 2008](#)].

El otro parámetro que fue estudiado es f_0 que da cuenta de la fuerza de cierre o apertura de los labios (controlado por el músculo dorsal syringealis) y es independiente de la dinámica misma de los labios. Las variaciones del valor de esta fuerza no modifican la estructura del diagrama de la figura 2.2 sino que generan corrimientos de la misma estructura a otros valores de los parámetros. En ese sentido, si el sistema está en la región 2 (oscilando) un aumento suficiente de f_0 puede generar que el diagrama se corra hacia la derecha haciendo que el sistema

no oscile más.

Vale la pena notar la presencia en este modelo de dos tipos de oscilaciones distintas generada mediante distintas bifurcaciones. Una de ellas capaz de generar sonidos tonales y otra capaz de generar sonidos con alto contenido espectral. Asimismo ambas son soluciones del mismo sistema y dependen de la región del espacio de parámetros (p_{sub}, k) en que se encuentre el mismo para que exista una u otra solución. Esta característica que tiene el modelo no estaba presente en los anteriormente desarrollados por el grupo ([Gardner et al., 2001, Laje et al., 2002]) y contempla justamente las principales características acústicas del canto del Diamante mandarín. En la siguiente sección se demuestra de manera cuantificada la pertinencia del modelo en términos de característica acústicas.

2.1.3. Cuantificación de características acústicas determinadas por la biomecánica.

El desarrollo del modelo condujo a analizar cuáles eran las característica acústicas que el mismo era capaz de reproducir. Para ello se procuró cuantificar dichas características (descritas en la figura 1.6) para estudiar cuán determinado está el canto por la biomecánica y comparar con los resultados de la síntesis del modelo.

Se determinaron dos parámetros relevantes para describir la acústica de las distintas sílabas del canto del Diamante mandarín. Ambos se obtienen de realizar un cómputo muy común al momento de analizar sonido que es la Transformada Rápida de Fourier (FFT). Este cómputo transforma la señal temporal, que es la onda de presión del sonido, en valores de frecuencia y energía para cada una de estas. El primer parámetro obtenido de cada segmento es la Frecuencia Fundamental Promedio (AFF, por su sigla en inglés) y fue calculado como el primer pico de la FFT por encima de algún valor umbral. El otro parámetro obtenido de cada segmento de sonido es el Espectro de Frecuencias Medio (MSF, por su sigla en inglés) y es la suma del producto de cada frecuencia por su energía correspondiente, dividido por la energía total del espectro [Sitt et al., 2008].

$$f_{MSF} = \sum_i \frac{\omega_i \epsilon_i}{E} \quad (2.6)$$

Sin embargo como el interés es hacer una cuantificación común que sirva para comparar todos los tramos de sonido de un canto, que como ya he dicho, presenta una gran variación de frecuencias fundamentales se construyó otro índice. Éste lleva el nombre de Índice de Contenido Espectral (SCI) e independiza al índice del valor de la frecuencia fundamental del segmento de canto estudiado [Sitt et al., 2008].

$$SCI = \frac{f_{MSF}}{f_{AFF}} \quad (2.7)$$

Se extrajo de 6 aves distintas una cantidad de 172 segmentos de sonido y computando ambos índices se ubicó cada segmento en el espacio determinado por (AFF, SCI) . El resultado se muestra en la figura 2.3 donde se puede observar que los 172 segmentos se ubican en una región acotada del espacio mostrando una clara relación entre ambos índices en un amplio rango de frecuencias. Asimismo el amplio rango de contenido espectral posible en las vocalizaciones y el hecho de que las sílabas de baja frecuencia sean generadas por pulsos en la dinámica de los labios [Jensen et al., 2007] sugiere fuertemente que la relación entre el contenido espectral y frecuencia fundamental que muestra la figura 2.3, es una restricción propia de la biomecánica. Para dar cuenta de ello se realizaron simulaciones con el modelo físico explorando el espacio de parámetros generando sonidos con distinta frecuencia fundamental. Se comparó entonces como se ubicaban en el espacio acústico definido por (AFF, SCI) los sonidos sintéticos respecto a los experimentales y se mostró que ambos se encuentran restringidos a una región del espacio (el resultado se muestra en la figura 2.3).

Por lo tanto a través de estos resultados se pudo mostrar que la biomecánica determina gran parte de las características acústicas del canto el Diamante mandarín. Esto resulta muy relevante al momento de analizar qué tipo de instrucciones neuronales requiere el sistema para generar canto, es decir las instrucciones pueden ser sencillas ya que gran parte de la complejidad acústica la determina el órgano fonador. Asimismo se comprobó que el modelo físico presentado en la sección anterior da cuenta de dichas restricciones acústicas incorporando los sonidos de baja frecuencia con alto contenido espectral (soluciones tipo pulsos que no eran consideradas en los modelos anteriores) [Sitt et al., 2008].

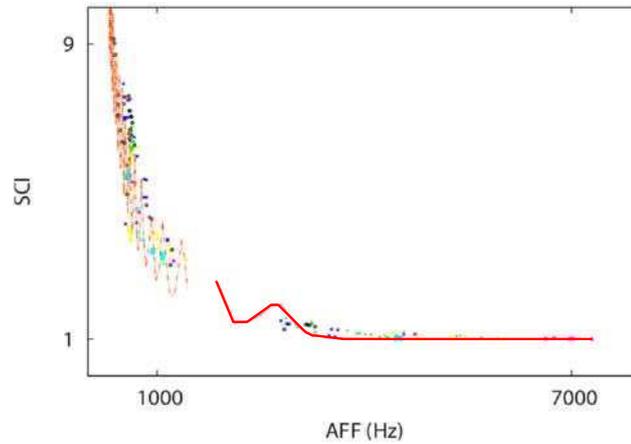


Figura 2.3: La figura muestra la relación entre contenido espectral (SCI) y frecuencia fundamental tanto para sonidos reales como para sonidos generados con el modelo. Cada sílaba de canto real analizada de las 172 empleadas en el análisis es representada en la figura a través de un punto y los distintos tipos de punto corresponden a distintas aves. La línea punteada representa la relación entre el SCI y la frecuencia fundamental para simulaciones numéricas donde con variaciones de la presión subsiringea (p_{sub}) se generaron sonidos de frecuencia menor a $1,5 \text{ kHz}$. La línea continua representa también simulaciones del modelo pero con variaciones de más parámetros barriendo el rango de alta frecuencia. Figura adaptada de [Sitt et al., 2008].

2.2. La forma normal

Una vez estudiado el modelo físico y su pertinencia desde el punto de vista acústico, los siguientes trabajos buscaron generar modelos que presenten el mismo escenario dinámico pero con menor costo computacional. Esto constituye un problema típico en el campo de los sistemas dinámicos donde se desarrollaron procedimientos algorítmicos que permiten hallar el sistema dinámico más sencillo que presente soluciones topológicamente equivalentes.

El sistema mostrado en la ecuación (2.5) presenta una característica particular además de todas las descritas en secciones anteriores. La dinámica cuenta con un punto específico que es donde la línea de la bifurcación de Hopf toca tangencialmente a la curva de la bifurcación de nodo-silla. Este punto es una bifurcación conocida como la bifurcación de Takens-Bogdanov y es una singularidad lineal donde ambas soluciones tienen autovalor 0. En el diagrama de la figura 2.2 se puede identificar el punto ya que se ubica en la juntura entre las regiones 1, 3, 4

y 5 [Guckenheimer and Holmes, 1997].

Las soluciones cercanas a singularidad de Takens-Bogdanov presentan una topología particular determinada por las características propias de la singularidad. Es decir, que cualquier sistema que presente una singularidad de este tipo en una región del espacio de parámetros lo suficientemente cercana a la singularidad va a respetar el escenario dinámico. El procedimiento típico entonces es llevar el sistema a la forma normal, siendo este el sistema más simple que mantiene el escenario dinámico. Para ello se realizan procedimientos algorítmicos típicos desarrollados en la literatura del campo de sistemas dinámicos [Wiggins, 2003].

El trabajo de J.Sitt se encargó de realizar una reducción a la forma normal del modelo físico presentado. Para ello se basó en los procedimientos típicos de cambios de coordenadas no lineales para eliminar los términos no resonantes y partiendo de desarrollos a tercer orden de la singularidad de Takens-Bogdanov se obtuvo la forma normal del modelo físico [Sitt et al., 2010], [Arrowsmith, 1990]:

$$\begin{aligned}\dot{x} &= y, \\ \dot{y} &= \gamma^2\alpha + \gamma^2\beta x + \gamma^2x^2 - \gamma xy - \gamma^2x^3 - \gamma x^2y.\end{aligned}\tag{2.8}$$

Donde α y β son funciones de la presión subsiringea y la actividad de los músculos vS respectivamente. Estas funciones fueron obtenidas mediante la transformación de las variables hecha para pasar del modelo físico a la forma normal detallado en [Sitt et al., 2010]. La constante γ representa un factor de escala temporal introducido para controlar que los valores de las variables mantengan su sentido fisiológico y den como resultado de su integración sonido comparable al medido experimentalmente. Este sistema tiene la particularidad de contar con la cantidad mínima de términos polinómicos no lineales y mantener el mismo escenario dinámico que el modelo físico. En particular este procedimiento logró eliminar términos donde se producía la división de polinomios que resultan costosos para el cómputo.

Se realizó un estudio sistemático de cómo se comportaba el modelo ante la variación de los parámetros. Se construyó por ende un diagrama de bifurcaciones para la forma normal en el espacio de parámetros transformado en (α, β) . Como era de esperar, por construcción las bifurcaciones existentes son las mismas por

lo tanto el tipo de soluciones posibles también. El diagrama de bifurcaciones se encuentra representado en la figura 2.4.

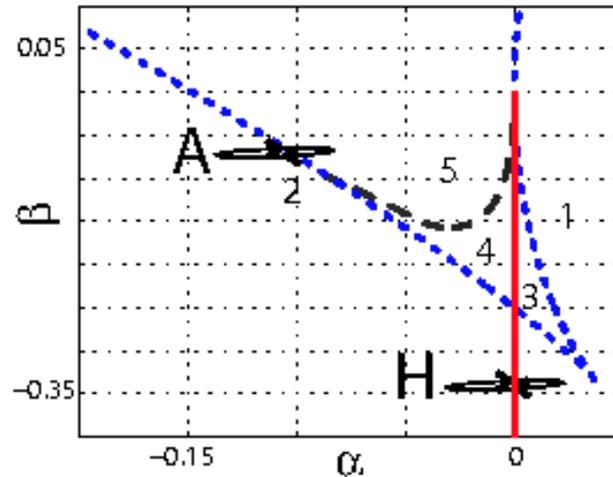


Figura 2.4: Se muestra el diagrama de bifurcaciones de la forma normal descrita por ecuación (2.8). El escenario dinámico planteado por el modelo físico se recupera en la forma normal, donde los parámetros de este diagrama son α y β . Éstos son producto de la reducción a la forma normal en torno a la singularidad de Takens-Bogdanov y están fuertemente relacionados con p_{sub} y k del modelo físico. Los números en el diagrama representan distintos regímenes dinámicos y están en correspondencia con los indicados en la figura 2.2. La línea negra punteada indica la bifurcación homoclínica, las líneas azules punteadas representan las bifurcaciones de nodo-silla y la línea roja, la bifurcación de Hopf.

Las bifurcaciones del modelo físico se mantienen y se indican con rojo la bifurcación de Hopf, con línea punteada azul la bifurcación de nodo-silla y con línea negra punteada la homoclínica. A su vez las transiciones relevantes para describir el canto del Diamante mandarín se identifican con una H y una A al igual que en la figura 2.2. Con H la bifurcación de Hopf que conduce a sonidos tonales y con A la bifurcación de SNILC (también conocida como de Andronov) con la que se logran las sílabas con alto contenido espectral. Asimismo los números que indican cada región se corresponden con las mismas regiones en la figura 2.2 y sus soluciones son las mostradas en los diagramas de fases correspondientes aquella figura.

Una vez realizada la reducción a la forma normal que optimiza los tiempos de

cómputo se planteó utilizar el modelo para sintetizar canto de Diamante mandarín realista. Para ello resultó necesario desarrollar avances en dos direcciones complementarias. Por un lado ajustar los parámetros del modelo para conocer cuáles combinaciones de los mismos resultan apropiadas para reproducir determinadas características acústicas; y por el otro, desarrollar un modelo para el tracto vocal superior que persiga obtener un timbre similar al del ave. Ambos avances fueron desarrollados en un trabajo del autor de esta tesis [Perl et al., 2011], el primero de ellos lo contaré en la sección siguiente, el segundo forma parte del siguiente capítulo.

2.3. Estudio e implementación de la forma normal

2.3.1. Ajuste factor de escala temporal

El primer parámetro a ajustar para utilizar la forma normal es el factor de escala temporal γ . Esto resulta indispensable para que el modelo recupere las características acústicas del canto en virtud de la relación encontrada entre frecuencia fundamental y SCI. Para ello extrajimos 28 segmentos de sonido correspondiente a 4 aves distintas cubriendo un amplio rango de frecuencias fundamentales y SCI. Luego a cada uno de ellos le computamos el valor de la frecuencia fundamental y de SCI correspondientes. Para cada valor de γ , realizamos segmentos de sonido sintéticos con el modelo tomando valores del espacio de parámetros (α, β) de una grilla de (100×100) . Para cada uno de los 28 segmentos de sonido experimental encontramos un sonido sintético que presente la frecuencia fundamental más próxima y calculamos entonces cual era el SCI que quedaba determinado para ese valor de frecuencia fundamental dado ese γ .

Sumando la diferencia al cuadrado entre el SCI experimental y el simulado para los 28 segmentos encontramos una forma de cuantificar cuán bueno resulta cada síntesis en función del γ .

$$\chi(\gamma)^2 = \sum_{28seg} (I_{SCI,exp} - I_{SCI,sin})^2 \quad (2.9)$$

2.3 Estudio e implementación de la forma normal

Donde $I_{SCI,exp}$ es el valor del SCI de los segmentos experimentales y $I_{SCI,sin}$ es el valor del SCI de los segmentos sintéticos. Graficamos por ende χ^2 en función de γ para un conjunto de 10 valores distintos, se puede observar el resultado en la figura 2.5. De esta forma encontramos el mínimo de ese gráfico para $\gamma = 24000$, asimismo repetimos el cálculo para cada uno de los 4 aves por separado obteniendo el mismo resultado. Por lo tanto consideramos que ese es el valor óptimo del factor de escala temporal para sintetizar canto de Diamante mandarín con el modelo propuesto. [Perl et al., 2011]

Vale aclarar que si bien aún no he contado nada del modelado de tracto vocal superior actuando como un filtro, las simulaciones de sonido se realizaron en este ajuste considerando un modelo de filtro promedio que será descrito en el siguiente capítulo. Asimismo todos los siguientes cálculos, no sólo se hicieron con un filtro promedio, sino que se realizaron también considerando el valor de $\gamma = 24000$ como resultado de este ajuste.

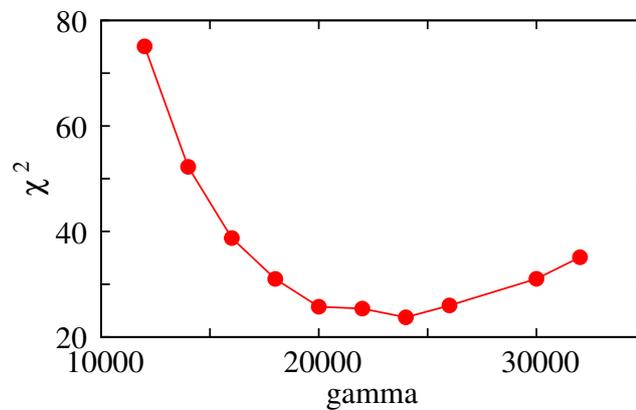


Figura 2.5: Se muestra el ajuste del parámetro γ de la ecuación (2.8). Para ello se computó para 28 segmentos de canto de 4 aves distintas el valor del SCI y de la frecuencia fundamental. Para cada valor de γ se generaron segmentos sintéticos buscando que la frecuencia fundamental sea lo más similar posible a la de los segmentos experimentales y luego se computó la diferencia entre el SCI del sintéticos y del experimental. Para cada γ se obtuvieron 28 diferencias entre SCI que se sumaron para dar un valor de χ^2 obteniendo como mínimo valor el obtenido para $\gamma = 24000$ como muestra la figura.

2.3.2. Curvas de iso-nivel para la frecuencia fundamental y el SCI

Hasta esta sección he contado por un lado, que el modelo físico se redujo a su forma normal quedando sólo 2 parámetros de los cuales depende el tipo de soluciones (α y β). Por el otro he descrito que la forma de caracterizar cuantitativamente al sonido del Diamante mandarín es en términos de la frecuencia fundamental y el SCI. Por ello nos propusimos conocer en detalle qué tipos de sonido era capaz de sintetizar el modelo en el espacio de parámetros (α, β) en términos de los dos parámetros acústicos.

Para conocer cómo se comporta el modelo en términos de los índices acústicos, realizamos simulaciones con el modelo generando sonidos sintéticos para una grilla de (100×100) de valores de los parámetros α y β . A cada segmento de sonido sintético le computamos el valor de la frecuencia fundamental y el SCI. Luego identificamos las curvas de iso-nivel para cada índice, es decir caminos dentro del espacio de parámetros (α, β) que mantuvieran la frecuencia fundamental o el SCI constante respectivamente. Como era de suponer los caminos para mantener el SCI constante no son los mismos que los que sostienen la frecuencia fundamental constante, es por ello que, como se muestra en la figura 2.6 en dos paneles, hay dos familias de curvas, una para cada índice acústico. Vale aclarar que como he contado en la sección anterior, el valor del factor de escala temporal se dejó fijo en el valor ajustado $\gamma = 24000$ para estas simulación, así como el filtrado de tracto superior se mantuvo en un filtrado promedio.

Para mayor comprensión mostramos las curvas de iso-nivel superpuestas con las bifurcaciones básicas de la forma normal (las curvas de la bifurcación de nodo-silla y Hopf). Nos permite ver cómo se comportan las soluciones en términos acústicos el las proximidades a las bifurcaciones. En la figura 2.6a) se muestran las curvas de iso-nivel para el SCI, la flecha que une los puntos c con d indica el sentido en la que decrece el valor de SCI. En el panel b) de la misma figura se muestran las curvas de iso-nivel para la frecuencia fundamental, la flecha que une los puntos e con f indica el sentido en el que crece la frecuencia fundamental. De esta forma se puede ver cómo cerca de la bifurcación de nodo-silla en ciclo límite,

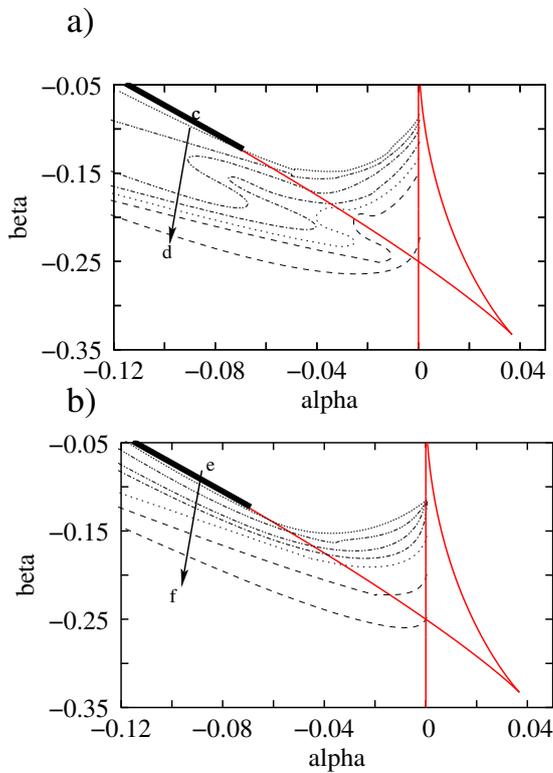


Figura 2.6: Las curvas de nivel para el SCI (a) y frecuencia fundamental (b) para el sistema dinámico definido por la forma normal dado por la ecuación (2.8). Para cada valor de los parámetros α y β se realizaron integraciones numéricas del modelo computando el SCI y la frecuencia fundamental del sonido generado. En el panel (a) se muestra un conjunto discreto de curvas de iso-nivel de SCI, la flecha que va de **c** hasta **d** indica el sentido en el que decrece el nivel de SCI de $SCI = 3,5$ hasta $SCI = 1,1$. En el panel (b) se muestra de manera análoga un conjunto discreto de curvas de iso-nivel de frecuencias fundamentales de sonidos generados por el modelo, la flecha que va desde **e** hasta **f** indica el sentido en el que crece el nivel de frecuencia fundamental de 750 Hz a 1900 Hz . Las líneas continuas rojas representan las bifurcaciones propias de la forma normal, la línea negra gruesa indica la zona donde sucede una bifurcación en presencia de ciclo límite (SNILC).

indicada con una línea gruesa negra en la figura 2.6, los sonidos presentan un alto contenido espectral (iso-nivel de $SCI = 3,5$) y baja frecuencia fundamental ya que en la bifurcación nacen con frecuencia cero (iso-nivel de $AFF = 750 Hz$) [Perl et al., 2011].

2.3.3. Reconstrucción de gestos motores

Una vez elegida la constante temporal γ de la forma normal y obtenidas las curvas de iso-nivel para una amplia región del espacio de parámetros (α, β) es posible reconstruir los parámetros α y β necesarios para sintetizar un canto de Diamante mandarín. Realizando este procedimiento algorítmico para sonidos secuenciales de un canto pudimos probar que los parámetros reconstruidos son una buena estimación de las series temporales de los parámetros fisiológicos empleados por el ave durante el canto. En [Perl et al., 2011] nos encargamos de obtener mediciones de dichas señales fisiológicas y las comparamos con las reconstruidas mediante un algoritmo de síntesis de canto. Asimismo es una forma de corroborar el modelo en términos, no de características acústicas, sino de gestos motores.

El algoritmo propuesto para la reconstrucción de gestos motores en búsqueda de la síntesis de un canto se inicia con la grabación del canto real empleando una frecuencia de muestreo de $44100 Hz$. Luego lo descomponemos en fragmentos de $20 ms$ que representa un tiempo lo suficientemente corto como para evitar grandes cambios en las señales fisiológicas y lo suficientemente largo como para calcular correctamente los dos elementos de nuestra descripción acústica. Por lo tanto para cada segmento calculamos los dos índices, la frecuencia fundamental y el SCI. Realizamos una búsqueda sobre el espacio de parámetros (α, β) hasta hallar los valores que den como resultado un sonido con índices acústicos lo más cercano posible a los experimentales. Para ello primero ubicamos la curva de iso-nivel de frecuencia fundamental más cercana a la computada experimentalmente y nos movemos sobre ella hasta dar con el valor de SCI que minimice la distancia con el SCI experimental. Una forma gráfica de comprender esta búsqueda es observando la figura 2.6, allí se puede ver como moverse sobre una curva de iso-nivel de frecuencia fundamental implica cruzar las curvas de iso-nivel de SCI, hasta dar con el adecuado. Este procedimiento define por ende unívocamente un

valor de α y de β que da como resultado el mejor ajuste en términos de los índices acústicos para cada segmento.

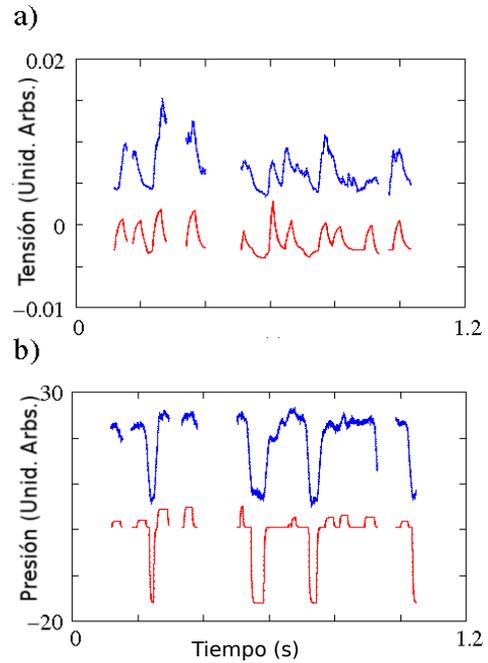
Los valores reconstruidos de α y β los empleamos para estimar la presión de los sacos aéreos y la actividad del músculo ventral (vS) respectivamente, mediante un cambio de signo multiplicándolos por un factor de escala [Sitt et al., 2010]. En la figura 2.7 en ambos paneles se muestran los parámetros α y β reconstruidos. Para generar cada uno de ellos en primer lugar reconstruimos un conjunto puntos discretos, uno cada 20 *ms*, tal como relaté en el párrafo anterior. Considerando que las señales fisiológicas son continuas en el tiempo y presentan variaciones suaves, generamos un trazo continuo uniendo los puntos. Para obtener los trazos continuos que muestra la figura 2.7 realizamos una integración del sistema:

$$\frac{du}{dt} = -\frac{1}{\tau}u + f(t) \quad (2.10)$$

donde $f(t)$ representa en cada caso a α y a β siendo funciones del tipo escalón donde mantienen constante el valor durante tramos de 20 *ms*. El parámetro τ representa un factor de escala temporal que fue elegido en $\tau = 150$ *ms*. Vale aclarar que los segmentos vacíos en la reconstrucción de los gestos corresponden a silencios producidos durante la expiración o bien segmentos aislados de canto con duración menor a los 20 *ms*.

En la figura 2.7 se muestra también en cada panel las mediciones directas sobre el ave cantando de la actividad muscular y la presión de los sacos aéreos. Ambas medidas requieren de equipo adecuado y son técnicas experimentales desarrolladas específicamente. Para el caso de la medición de actividad del músculo vS, se realiza una cirugía donde se insertan en el músculo pequeños electrodos. Estos consisten en cables de acero inoxidable aislado, adheridos al tejido muscular con una micro-gota de pegamento. Una vez insertados los electrodos correctamente se cierra al zona abierta, se retiran los cables y dirigen hacia la parte dorsal del ave [Goller and Suthers, 1996a]. Para el caso de la presión en los sacos aéreos, la medida se realiza insertando una cánula en saco aéreo abdominal conectada en el otro extremo a un transductor de presión [Suthers and Margoliash, 2002]. En la figura 2.7 se muestra mediciones simultaneas de los valores fisiológicos por encima de cada reconstrucción mediante el modelo.

Figura 2.7: Los parámetros reconstruidos a partir de la implementación del algoritmo descrito comparados con las señales fisiológicas medidas durante el canto. En el panel a) se observa la tensión del músculo vS medida y los valores reconstruidos del parámetro $|\beta|$ en azul y rojo respectivamente. En el panel b) se muestra la presión de los sacos aéreos medida durante el canto y los valores reconstruidos del parámetro $|\alpha|$ en azul y rojo respectivamente.



Con las medidas fisiológicas y las series temporales de las reconstrucciones para todo un canto buscamos obtener un valor que de cuenta de cuán buena estimación es la reconstrucción de la medida experimental. Para ello computamos la correlación $C(exp, sin)$ entre ambas señales temporales, los valores obtenidos para cada medida fueron, $C = 0,72$ para actividad muscular y $C = 0,91$ para la presión de los sacos aéreos. Si bien para el caso de la actividad muscular es menor, son valores altos de correlación por lo que este resultado conduce a probar la pertinencia del modelo en términos de los gestos motores. Es decir, que las variables seleccionadas para controlar el modelo, la actividad muscular y la presión de los sacos, son efectivamente responsables de dicho control en la biomecánica del ave [Perl et al., 2011].

Analizando la reconstrucción de la actividad muscular del vS podemos realizar varias observaciones. Por un lado como ya he dicho la siringe es una estructura bipartita y al realizar un canto el ave emplea ambas fuentes sonoras. Nosotros en la figura 2.7 mostramos sólo la actividad de la estructura de la derecha, sin embargo computamos la correlación de nuestra reconstrucción con ambos la-

dos dando significativamente mayor la correlación con la fuente de la derecha ($C(\beta, vS_{derecha}) = 0,72 > C(\beta, vS_{izquierda}) = 0,56$). Sin embargo nuestro algoritmo de reconstrucción tiene como punto de partida el canto del ave que involucra la actividad simultanea de ambas fuentes. La alta correlación de β con la actividad con el vS derecho se debe probablemente a que en efecto el ave al cantar utiliza la fuente de la derecha para todas las sílabas, las de baja y alta frecuencia. No así la fuente izquierda que la utiliza solo para las sílabas de alta frecuencia y nunca está activa sola [Goller and Cooper, 2004].

Con los gestos motores reconstruidos mediante el algoritmo descrito, podemos ir tras una síntesis del canto. Es allí donde desarrollamos trabajos que modelan el tracto vocal superior en la búsqueda de mejoras tímbrica. Esa búsqueda de generar canto sintético realista será el eje del capítulo siguiente.

Capítulo 3

Una síntesis realista

Hasta este capítulo conté el desarrollo y estudio de los modelos dinámicos que describen el funcionamiento de la siringe. Sin embargo, en el capítulo introductorio describí la anatomía del Diamante mandarín donde entre el órgano fonador y el ambiente se ubica el tracto vocal superior que afecta al sonido final que emite el ave. Asimismo indiqué la existencia de dos teorías que describen las implicancias del tracto vocal superior sobre el sonido, en este trabajo voy a considerar la teoría de fuente-filtro desacoplados. Eso implica que por un lado existe la dinámica de la siringe controlada por los parámetros α y β y por el otro la del tracto vocal superior que actúa filtrando al sonido cuya dinámica es independiente y no afecta a la de la siringe. La primera determina, como hemos visto en el capítulo anterior, las características acústicas esenciales, el segundo se encarga de dar el filtrado acercando las características tímbricas del canto generado con el modelo a las medidas experimentalmente.

En la búsqueda de generar síntesis realista se desarrollaron distintos modelos del tracto vocal superior dentro de la teoría de fuente-filtro desacopladas. Nuevamente, la filosofía de estos modelos propuestos por el Laboratorio de Sistemas Dinámicos es generar modelos minimales logrando establecer jerarquías en los elementos involucrados en el sistema. Es decir, cuál es el menor número de variables y parámetros capaces de reproducir de manera realista y con bajo costo computacional el comportamiento complejo estudiado.

3.1. Modelos de trato vocal superior

La presencia de un tracto vocal superior independiente de la siringe aparece, como he dicho, como un filtro al sonido. Eso implica que las características del tracto no generan nuevos sonidos ni determinan propiedades intrínsecas a la generación del mismo, sino que resaltan o atenúan algunas secciones del espectro sonoro. En esa dirección en diversos trabajos se han desarrollado modelos que filtran el sonido mediante representaciones simples del tracto vocal superior. Los más relevantes en la literatura son dos tipos de modelos, uno que modela a la tráquea y al pico como tubos de distinta área y longitud (típica en modelos de habla humana) [Gardner et al., 2001] y otro que es más específico para Diamante mandarín que modela la anatomía del tracto vocal superior como un tubo, una cavidad resonante y un extremo irradiante [Perl et al., 2011, Perl et al., 2012].

3.1.1. Modelos de tubos

Uno de los primeros modelos que surgió en el área de canto de aves para dar las características acústicas finales a los sonidos sintéticos fue el que le asigna tubos a las distintas secciones del tracto. En particular el primer trabajo que incorporó esta idea para la síntesis del canto de aves (Canarios en este caso) fue el de T. Gardner [Gardner et al., 2001] y lo hizo tomando las ideas de Titze quien lo había planteado para el filtrado del habla en humanos [Titze, 1994].

En dicho trabajo consideraron al tracto vocal superior como la tráquea y el pico que fueron modelados como dos tubos acoplados de área (A_1 y A_2) y longitud distinta (L_1 y L_2). La forma que emplearon para comunicar la siringe con la tráquea fue considerar que a la entrada de ésta existe una presión $P_i(t)$ que es proporcional a la derivada del flujo en esa posición. Esta presión genera una onda que se ve afectada por la geometría de los tubos reflejándose y transmitiéndose parcialmente en cada interfase. La cantidad de energía que se refleja en la interfase es proporcional al coeficiente $r_{1,2} = (A_1 - A_2)/(A_1 + A_2)$, la energía transmitida entonces estaría determinada por $t_{1,2} = 1 - r_{1,2}$. De esta forma implementaron un modelo de rebotes en tubos con ondas propagándose en ambas direcciones. Llamaron $a(t)/c(t)$ a las ondas propagándose hacia adelante en el primer y segundo tubo respectivamente y $b(t)/d(t)$ a las que se propagan hacia atrás para escribir

el modelo:

$$\begin{aligned}
 a(t) &= P_i(t) - b(t - \tau_1) \\
 b(t) &= r_{1,2}a(t - \tau_1) + t_{1,2}d(t - \tau_2) \\
 c(t) &= t_{1,2}a(t - \tau_1) + r_{1,2}d(t - \tau_2) \\
 d(t) &= \alpha c(t - \tau_2)
 \end{aligned}
 \tag{3.1}$$

donde α es la reflexión en el segundo tubo en la interfase con el ambiente y τ_i es lo que tarda en recorrer el tubo de longitud L_i . Es importante observar que este es un conjunto de ecuaciones acopladas con retardos, pues se evalúa la misma función que se quiere conocer en otros tiempos, lo que implica un sistema de dimensión infinita. Asimismo en este trabajo no se consideró la dinámica del pico (sería la presencia de $\alpha(t)$) y por simplicidad se dejó fija la apertura del mismo sin tener mayores consecuencias tímbricas [Gardner et al., 2001].

Los siguientes trabajos, como el de J. Sitt buscaron una versión más sencilla y fueron tras un modelo de un único tubo de longitud L y coeficiente de reflexión r . En este caso la presión a la entrada de la tráquea P_i tiene dos componentes, la directa del flujo proveniente de la fuente sonora y el rebote en el extremo del tubo:

$$P_i(t) = p_f(t) - rP_i(t - \frac{2L}{\nu})
 \tag{3.2}$$

donde el primer término de la derecha $p_f(t)$ representa la presión en la entrada de la tráquea producto del flujo de aire modulado por la siringe. Se lo puede escribir como $\lambda x(t)$ donde λ es proporcional a la velocidad media del flujo y $x(t)$ es la solución del sistema dinámico propuesto para la fuente [Arneodo and Mindlin, 2009]. El segundo término es la presión en la entrada de la tráquea pero generada por la onda que vuelve luego de haber rebotado en el extremo del tubo viajando a velocidad del sonido ν . En este caso también se trata de una ecuación con retardos, pues para conocer $P_i(t)$ es necesario usar $P_i(t - 2L/\nu)$ que es la misma función evaluada en otro tiempo, lo que implica nuevamente un sistema de infinitas dimensiones.

Otros modelos utilizan tres tubos que dan cuenta de la tráquea, la OEC y el pico y su implementación es la misma que la de de los modelos comentados. Primero se controla al modelo de siringe con los parámetros adecuados para generar soluciones oscilatorias y luego esa salida se introduce al inicio de sistema de

tubos para generar el filtrado sonoro [Laje et al., 2002, Mindlin et al., 2003].

En todos los casos de modelos de tubos, el filtro actúa como algo estático que no cambia con el tiempo. Es decir no son modelos capaces de representar modificaciones del tracto que acompañen los cambios acústicos a medida que el ave genera el canto. En la siguiente sección contaré el desarrollo de un modelo de tracto vocal que además de ser más específicos en su modelado anatómico, tienen la capacidad de ser dinámico y operacional, y por ende variable durante el canto.

3.1.2. Modelo con OEC

En trabajos posteriores al desarrollo de los modelos de tubos narrados en la sección anterior se realizaron nuevas observaciones sobre el funcionamiento del tracto vocal. Uno de los más relevantes es el de N. Fletcher y co-autores, ellos notaron que algunas aves son capaces de modificar la geometría de su tracto vocal de forma coordinada con las modulaciones de frecuencia del canto [Fletcher et al., 2006a]. Asimismo conjeturan y luego miden que cambios temporales en partes del tracto son compatibles con cambios en algunas características acústicas. De esta manera realizaron un modelo más realista del tracto vocal e identificaron que la cavidad orofaríngea-esofágica (OEC) es dinámicamente ajustada por el ave con el fin de enfatizar la frecuencia que varía en el tiempo [Fletcher et al., 2006b, Ohms et al., 2010].

Por lo tanto nosotros desarrollamos un modelo que recuperó esa propiedad identificada en los citados trabajos. En esa dirección tratamos al filtro como un sistema dinámico aproximando la tráquea por un tubo y la OEC por un resonador de Helmholtz cuyo cuello de entrada es la glotis y apertura hacia el ambiente, el pico [Perl et al., 2011, Perl et al., 2012]. En la parte superior de la figura 3.1 se muestra el esquema del modelado del tracto.

La descripción del tracto la estructuramos en dos etapas, la primera consiste en un tubo abierto-cerrado en sus extremos cuya salida fuerza a una segunda etapa que consiste en una constricción, un resonador de Helmholtz y una abertura irradiante. El modelado formal del tubo, al igual que los relatados en la sección anterior, recibe en la entrada una presión proporcional al flujo del aire modulado por la dinámica de la siringe $P_i(t)$. El resultado del sonido a través de esta

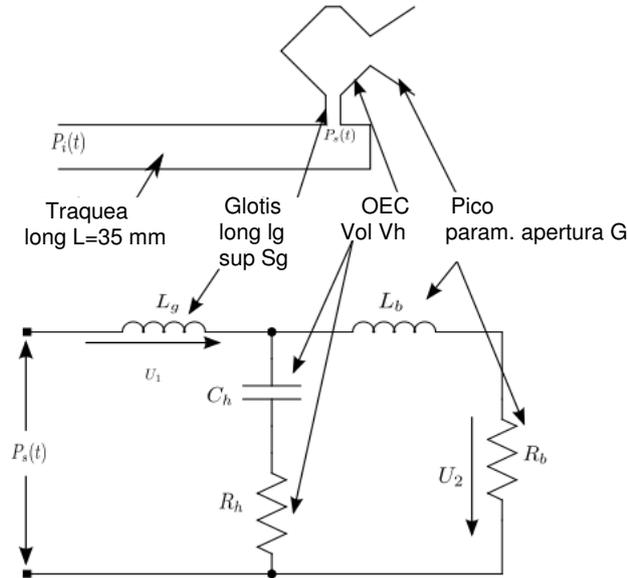


Figura 3.1: El modelo incluye una válvula, cuya dinámica está descrita por el modelo detallado en el capítulo anterior, y un tracto vocal pasivo que se ilustra en esta figura que filtra el sonido generado en la fuente. El modelo del tracto vocal superior consiste en una tráquea, una cavidad que representa a la cavidad oroeso-faríngea (OEC), una glotis y un pico. El circuito ilustrado en la parte inferior de la figura es el análogo eléctrico del filtro que describe la dinámica de un resonador de Helmholtz (OEC) en paralelo con la apertura hacia el exterior (el pico).

etapa es un filtrado que enfatiza algunas frecuencias en particular vinculadas con la longitud del tubo. La ecuación (3.2) representa de forma adecuada el efecto del tubo (la tráquea) sobre el sonido considerando un rebote y permite realizar implementaciones computacionales de esta etapa del filtro. Luego la presión que emerge de la tráquea $P_s(t)$ es la que fuerza al sistema acústico glotis-OEC-pico, cuya expresión es $P_s(t) = (1 - r)P_i(t - L/\nu)$, con r índice de reflexión, L longitud del tubo y ν la velocidad del sonido en el mismo. En la dirección de conocer cómo evoluciona temporalmente el sonido al atravesar un sistema acústico variable escribimos un sistema dinámico de baja dimensión que lo representa.

Es común en sistemas acústicos escribir modelos análogos eléctricos para de-

scribir al sistema de filtros. La presión acústica es representada por un potencial eléctrico y el volumen de flujo por una corriente eléctrica [Fletcher, 1992]. Asimismo una cavidad resonante conectada a través de un cuello es modelada en el análogo eléctrico como una inductancia L_g en serie con un capacitor C_h . Si además se consideran las pérdidas en el resonador, se agrega una resistencia R_h en serie a los elementos anteriores. El resonador de Helmholtz se encuentra abierto al ambiente (a través del pico), el circuito análogo que lo representa consiste en una inductancia L_g en serie con una resistencia R_b que se conectan en paralelo al capacitor que representa al resonador [Fletcher et al., 2006a]. En la figura 3.1 se muestra (abajo) el circuito equivalente eléctrico que representa el sistema acústico que modela el tracto vocal superior (arriba).

Las ecuaciones del equivalente eléctrico de la parte post-traqueal del filtro se leen:

$$\begin{aligned}
 \frac{di_1}{dt} &= \Omega_1, \\
 \frac{d\Omega_1}{dt} &= -\frac{1}{L_g C} i_1 - R_h \left(\frac{1}{L_b} + \frac{1}{L_g} \right) \Omega_1 + i_3 \left(\frac{1}{L_g C} - \frac{R_b R_h}{L_b L_g} \right) + \frac{1}{L_g} \frac{dV_{ext}}{dt} + \frac{R_h}{L_g L_b} V_{ext}, \\
 \frac{di_3}{dt} &= -\frac{L_g}{L_b} \Omega_1 - \frac{R_b}{L_b} i_3 + \frac{1}{L_b} V_{ext},
 \end{aligned} \tag{3.3}$$

donde Ω_1 representa la derivada temporal de i_1 y la relación entre los componentes del análogo eléctrico y los elementos acústicos se muestran en la tabla 3.1 [Kinsler et al., 1982].

Impedancia	Inertancia
L	$M = \frac{\rho_0 l_a}{S_a}$
Resistencia	Resistencia
R	$R = \frac{\rho_0 c k^2}{2\pi}$
Capacitancia	Elasticidad
C	$C = \frac{V_h}{\rho_0 c^2}$

Tabla 3.1: Análogo acústico y eléctrico

En dicha tabla ρ_0 es la densidad del aire y c la velocidad del sonido. El volumen de la cavidad h es V_h e indica el volumen en nuestro caso del resonador de Helmholtz que representa la OEC. La longitud y el área del elemento a son l_a y S_a respectivamente, en nuestro modelo consideramos dichas magnitudes para el pico (b) y para la glotis (g). En el caso de la estimación de la constricción realizada por el pico que da lugar a la aparición de L_b no utilizamos la equivalencia planteada en la tabla 3.1 sino una aproximación específica que considera también el aporte de la lengua quedando entonces para el término inertivo $L_b = \rho_0/G$ donde G es un parámetro de apertura [Fletcher et al., 2006a]. Usando las equivalencias mostradas en la tabla 3.1 y los rangos de valores fisiológicos estimados de cada uno de los parámetros acústicos en [Fletcher et al., 2006a], pudimos derivar valores posibles para los parámetros de nuestro sistema dinámico equivalente.

Aproximando la tráquea como un tubo abierto-cerrado en sus extremos de 3,5 cm [Daley and Goller, 2004] pudimos dar cuenta de la presencia ubicua de frecuencias enfatizadas en $f_1 = 2,5 \text{ KHz}$ y $f_2 = 7,5 \text{ KHz}$ en el canto del Diamante mandarín. Sin embargo es importante marcar que en la mayoría de los casos la primera resonancia de la OEC es la frecuencia dominante en el espectro. El valor de esta frecuencia de resonancia depende fuertemente del volumen de la OEC y del área de la glotis (C_h y L_g son los parámetros del equivalente eléctrico respectivamente). En la figura 3.2 se muestran cuatro paneles con las características espectrales del filtro, correspondientes a distintas combinaciones de los parámetros. La frecuencia entre $f_1 = 2,5 \text{ KHz}$ y $f_2 = 7,5 \text{ KHz}$ en cada panel corresponde a la primera resonancia del resonador de Helmholtz. Se puede observar que para un valor fijo de volumen de la OEC, al disminuir la apertura de la glotis (aumentar la impedancia glotal) baja el valor de la frecuencia característica (figura 3.2 a) y b)). Del mismo modo a valor constante de la apertura glotal aumentando el volumen de la OEC, como era de esperar, hace decrece el valor de la frecuencia característica. Asimismo el valor R_h controla la amplitud de la resonancia de la OEC y por lo tanto el peso relativo que tiene respecto a las frecuencias enfatizadas por la tráquea [Perl et al., 2011].

Por otro lado también se realizaron simulaciones numéricas variando los parámetros que dan cuenta de la constricción que genera el pico y de las pérdidas que suceden en el mismo. Notamos que los cambios en las características espectrales

del filtrado eran muy sutiles. Al variar el parámetro que da cuenta de las pérdidas R_b , como era de esperar, provocó una disminución en la amplitud total del sonido resultante. Al variar L_b , se observaron pequeñas variaciones en el espectro en la región de altas frecuencias. De esta manera el conjunto de parámetros que empleamos para controlar las características espectrales del filtro son L_g , C_h y R_h .

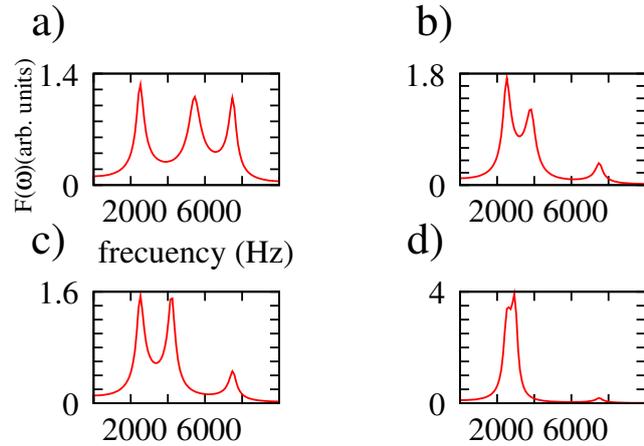


Figura 3.2: En nuestro modelo el tracto vocal superior consiste en un tubo abierto-cerrado de 3,5 cm seguido de un resonador de Helmholtz. En el rango mostrado el tubo aporta dos resonancias $f_1 = 2,5 \text{ KHz}$ y $f_2 = 7,5 \text{ KHz}$. La tercer resonancia en cada panel está determinada por la primer frecuencia característica del resonador de Helmholtz, dependiendo de los parámetros de la cavidad. Para cada panel (a) $C_h = 1,5 \times 10^{-8} / 350 \frac{\text{m}^3}{\text{Pa}}$ y $L_g = 20 \frac{\text{Pa} \cdot \text{s}^2}{\text{m}^3}$; (b) $C_h = 1,5 \times 10^{-8} / 350 \frac{\text{m}^3}{\text{Pa}}$ y $L_g = 40 \frac{\text{Pa} \cdot \text{s}^2}{\text{m}^3}$; (c) $C_h = 2,5 \times 10^{-8} / 350 \frac{\text{m}^3}{\text{Pa}}$ y $L_g = 20 \frac{\text{Pa} \cdot \text{s}^2}{\text{m}^3}$; (d) $C_h = 2,5 \times 10^{-8} / 350 \frac{\text{m}^3}{\text{Pa}}$ y $L_g = 40 \frac{\text{Pa} \cdot \text{s}^2}{\text{m}^3}$; Los demás parámetros se mantuvieron fijos en los cuatro paneles: $R_h = 1000 \frac{\text{Pa} \cdot \text{s}}{\text{m}^3}$, $L_b = 1000 \frac{\text{Pa} \cdot \text{s}^2}{\text{m}^3}$, $R_b = 100000 \frac{\text{Pa} \cdot \text{s}}{\text{m}^3}$. Los parámetros tienen unidades de los elementos del equivalente acústico [Perl et al., 2011].

Así mostramos que con el modelo dinámico de filtrado es posible ir más allá que generar canto sintético ajustando las frecuencias fundamentales y el SCI. Es decir, colocando ambos modelos juntos controlando cada uno con sus parámetros correspondientes podemos sintetizar sonido realista con características tímbricas adecuadas. Sin embargo a la hora de pensar la pertinencia de este tipo de modelados surgen muchas preguntas: *Cuáles son las características tímbricas relevantes?*, *Serán las mismas características que nosotros identificamos las*

que le importan al ave?, Cuál modelo de filtrado es adecuado entonces? Como responderá el ave ante un canto sintético?. Todas estas preguntas fueron las que perseguimos responder en parte del trabajo cuyo desarrollo contaré en lo que queda del capítulo.

3.2. Pertinencia de los modelos en términos de respuesta neuronal

En la dirección de responder parte de las preguntas planteadas en la sección anterior, fue que dejamos de lado las distintas medidas computacionales para ir tras métricas experimentales en neuronas. Es decir, anteriormente comprobábamos la pertinencia del modelado a través de parámetros acústicos o bien a través de comparación de gestos motores. En este trabajo fuimos tras probar la pertinencia en términos de respuesta neuronal. Para ello nos basamos en determinadas características que presentan algunos núcleos neuronales involucrados en el sistema del canto como ser la selectividad y el fenómeno de réplica. Asimismo contamos con la colaboración de Ana Amador que durante su post-doctorado en la Universidad de Chicago incorporó técnicas de medición de actividad neuronal en diversos núcleos neuronales del sistema del canto.

3.2.1. Selectividad neuronal

Como he contado en la sección introductoria existen diversos núcleos neuronales involucrados en el proceso de aprendizaje y generación del canto en aves. En particular existe el núcleo conocido como área L que vincula la Vía auditiva con el HVc que forma parte de la Vía motora. Es por ello que se realizaron diversos estudios para ver qué tipo de estímulo auditivo era transmitido hasta el HVc a propósito de comprender cómo y qué tipo de información codifica dicho núcleo.

En uno de los primeros estudios en esa dirección Katz y Gurney reportaron que en mediciones intra-celulares en aves anestesiadas, muchas neuronas del HVc respondían débilmente a ráfagas de ruido como estímulo auditivo [Katz and Gurney, 1981]. Luego McCasland y Konishi trabajaron con aves despiertas y mediante mediciones de muchas unidades de neuronas simultáneas (“multiunits”)

de HVc observaron que presentaban actividad al escuchar grabaciones de canto conoespecífico y del canto propio (BOS por su sigla en inglés “bird’s own song”), y no así al canto reproducido en reversa cuya actividad era significativamente menor [McCasland and Konishi, 1981]. En trabajos en gorriones de corona blanca (*Zonotrichia leucophrys*) D. Margoliash fue más allá e identificó mediante manipulación del canto propio del ave cuáles eran las características acústicas a las cuales las neuronas del HVc presentaban actividad. De esta manera reportó que la selectividad de las neuronas de HVc se vincula con características tanto espectrales como temporales, que no depende sólo de un valor de frecuencia fundamental sino de sonidos apareados, transiciones o secuencias del canto. [Margoliash, 1983]. En particular estas neuronas son altamente selectivas en su respuesta a características complejas del canto propio del ave (BOS) y se encuentran embebidas en una población más grande de neuronas de HVc que presentan menor selectividad al canto propio [Margoliash, 1986].

A partir de las observaciones de los trabajos anteriores se comenzó entonces a estudiar la selectividad de determinadas neuronas en HVc al canto propio. En [Margoliash, 1986], también en gorriones de corona blanca, el autor aportó varias de las pruebas en esa dirección. Mediante mediciones de muchas neuronas en simultáneo en HVc estudió la respuesta neuronal diferenciada ante distintos cantos. De esta forma probó la existencia de algunas neuronas específicas que respondían significativamente más al canto propio que al conoespecífico incluyendo cantos similares y el canto del tutor. Asimismo mostró que la selectividad de este grupo de neuronas era estable en aves adultas y que modificaciones en la canción generaban respuestas menores. Otra observación relevante que fue reportada en este trabajo es que el área L no presenta selectividad neuronal al canto propio, por lo que dice es un indicio de que es algo propio generado en la plasticidad de HVc. Asimismo sugiere que estas neuronas altamente selectivas de HVc entonces podrían encargarse de “filtrar” el canto propio respecto de otros cantos percibidos.

En otro trabajo posterior D. Margoliash y E. Fortune buscaron extender la selectividad a otras especies y se preguntaron justamente si el Diamante mandarín mantenía la cualidad observada en gorriones de corona blanca. Efectivamente, midiendo en machos de Diamante mandarín adulto anestesiados actividad neuronal de unidades únicas en HVc y encontraron que las mismas presentaban una

alta selectividad a combinaciones de fragmentos del canto propio [Margoliash and Fortune, 1992].

3.2.2. Fenómeno de réplica

La otra característica importante en la cual basamos nuestras investigaciones es el fenómeno de réplica. Éste consiste en la capacidad de las neuronas de algunos de núcleos de generar el mismo patrón de actividad cuando elicitan respuesta al ser estimuladas con grabaciones de canto propio que cuando ejecutan su canto. En el trabajo [Dave and Margoliash, 2000] A. Dave y D. Margoliash trabajando con Diamante mandarín identificaron y demostraron este fenómeno realizando medidas de neuronas únicas en el núcleo RA. Es decir, confirmaron en sus mediciones que la selectividad neuronal se conserva en el núcleo RA en Diamante mandarín y que su estructura de disparo representa un patrón temporal vinculado al canto. Asimismo observaron que la actividad espontánea de estas neuronas durante el sueño coincide también con la actividad durante el canto, es decir que replica la estructura y los tiempos de disparo espontáneamente mientras duerme [Dave and Margoliash, 2000]. Esto sugiere que durante el día se genera un patrón de respuesta que luego es replicado durante el dormir cuando no se está comportando como mecanismo de aprendizaje y generación de un canto propio [Shank and Margoliash, 2008].

Un trabajo posterior que estudió el fenómeno de réplica en HVc fue el de J. Prather y co-autores. En [Prather et al., 2008] notaron estudiando Gorriones pantaneros (*Melospiza georgiana*) en neuronas proyectoras de HVc al área X que presentan la misma estructura temporal cuando el ave ejecuta el canto y que cuando lo escucha. A diferencia de los estudios relatados en el párrafo anterior en este caso el ave presenta el fenómeno de réplica despierta y comportándose libremente. Eso significa que estas neuronas que presentan una respuesta precisa al estímulo auditivo y la generación vocal, responden de igual manera cuando están generando el canto como cuando lo escuchan.

En Diamante mandarín, existe una notable dependencia del estado del ave (despierta o dormida) al fenómeno réplica neuronal, es decir si el ave está despierto no presenta este fenómeno [Dave and Margoliash, 2000] y se asocia fuertemente

con el aprendizaje del canto [Shank and Margoliash, 2008]. Asimismo la mayor respuesta y mayor selectividad auditiva en HVC fue medida en aves dormidas. Es por ello que nos propusimos contrastar nuestros modelos empleando mediciones neuronales en Diamante mandarín dormido aprovechando que la respuesta y selectividad es mayor en ese estado.

3.2.3. Comprobación de los modelos

Para chequear la pertinencia de los modelos comparamos la respuesta neuronal HVC en aves dormidas al reproducirles canto sintético con la obtenida al reproducirle canto propio. Como he descrito en la sección anterior en aves dormidas, el sistema del canto presenta neuronas excepcionalmente selectivas y que resulta muy difícil elicitarse respuesta en ellas. Es por ello que elegimos esta métrica de comparación como aquella que resulta significativa para el ave.

Para ello el modo de trabajo consistió en dos equipos trabajando de forma complementaria y en simultáneo. Por un lado en el Laboratorio de Sistemas Dinámicos nos encargamos de modelar cantos de Diamante mandarín cuyas grabaciones eran enviadas desde el Department of Organismal Biology and Anatomy en la Universidad de Chicago. Por el otro, en Chicago, A. Amador grababa cantos de animales de la colonia propia de la Universidad, realizaba las cirugías pertinentes y recolectaba los datos de electrofisiología.

El protocolo empleado para sintetizar cantos fue descrito en el capítulo anterior, donde a partir de grabaciones de los mismos reconstruimos mediante la forma normal del modelo de siringe los gestos motores ($\alpha(t)$ y $\beta(t)$). Luego con esos gestos obtenidos integramos las ecuaciones del modelo para obtener la presión a la salida de la fuente (proporcional a $x(t)$) que es insertada en el modelo de tracto vocal superior para filtrar acústicamente el sonido. El resultado es un canto sintético realista cuya pertinencia fuimos a probar [Perl et al., 2011, Perl et al., 2012].

El protocolo empleado para las mediciones de actividad neuronal está descrito en detalle en [Amador et al., 2013]. Las cirugías fueron realizadas por A. Amador en concordancia con los protocolos aprobados por el Comité de Cuidado y Uso Animal de la Universidad de Chicago y consistieron en la inserción de

un “pin” en la cabeza de las aves para fijarlas en un soporte estereotáxico. Luego se mantuvieron en cajas acústicamente aisladas con un ciclo de luz invertido de $16 - 8$ *hs* y sobre el ave fija se practicaron mediciones extracelulares de neuronas de HVC. Las mediciones fueron luego post-procesadas por un algoritmo de separación de espigas desarrollado en la U. de Chicago que permite diferenciar los tiempos de la actividad de cada neurona presente en la grabación. Para la presentación de los estímulos Amador editó cantos incluyendo dos o tres repeticiones del motivo para el canto propio (BOS), el canto sintético (mBOS por su sigla en inglés), el canto de un conespecífico y el canto en reverso. Los cuatro estímulos fueron presentados de forma aleatoria con intervalos entre estímulo de 7 ± 1 *s* un total de 20 – 25 veces cada uno.

La respuesta neuronal a cada estímulo fue cuantificada en términos de un índice conocido como el valor Z [Fiete et al., 2004]. La respuesta al modelo del canto propio (mBOS), al canto conespecífico (CON) y al reverso (REV) fue normalizada al valor Z del BOS, promediada a través de las neuronas y reportada como el promedio normalizado \pm s.e.m. Vale aclarar que el protocolo de presentación de estímulos incluye al conespecífico y al reverso como comprobación en cada medición de la selectividad de la neurona medida. Se consideró que una neurona funcionaba selectivamente al canto propio a aquella que presentaba respuesta al canto propio (BOS) significativamente mayor que al canto conespecífico y al reverso.

Comprobación modelo de tubos

Una vez desarrollado el protocolo, fuimos a realizar las pruebas con los cantos sintéticos sobre las aves. Como he contado existen dos modelos de tracto vocal superior, el de tubos y el que involucra a una cavidad resonante (OEC). En primera instancia, fuimos a probar la pertinencia del **Modelo de Tubos**, es decir modelamos los gestos motores con la forma normal, generamos el canto que luego filtramos con un tubo y finalmente se realizó en Chicago la comprobación neuronal. El resultado del proceso de síntesis se observa en la figura 3.3. Vale la pena notar que en términos de duración de sílabas, frecuencias fundamentales y la audición de los modeladores la síntesis realmente coincide con el canto grabado.

Sin embargo el resultado del experimento sobre el ave fue categórico, la síntesis no elicó respuesta neuronal. La conclusión que obtuvimos de este resultado negativo fue que existen algunas sutilezas tímbricas que nuestro modelo minimal de una fuente seguida por un tubo evidentemente no recuperó. Por lo tanto en el afán de realizar modelos minimales del órgano fonador aviar, este resultado nos pone en evidencia que la reducción propuesta era insuficiente para generar síntesis realista en términos de respuesta neuronal.

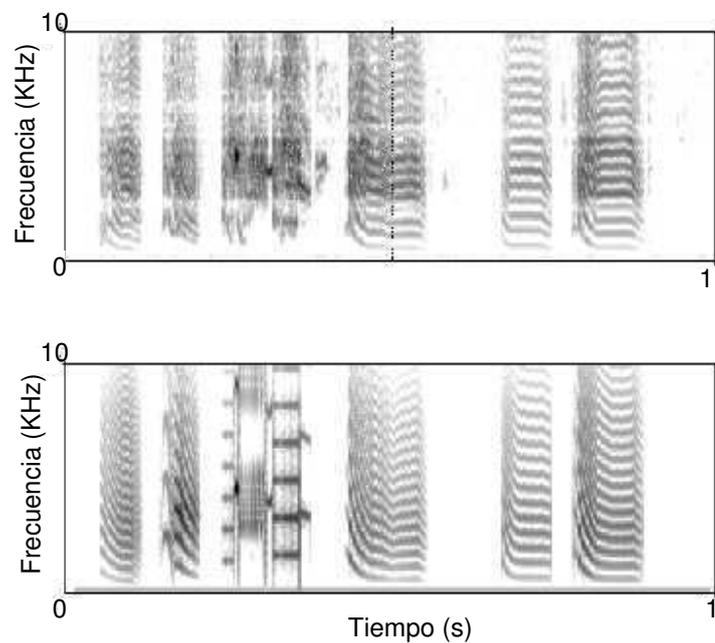


Figura 3.3: La figura muestra dos sonogramas, el superior corresponde a la grabación de un canto real de Diamante mandarín, el inferior corresponde al canto obtenido sintéticamente del mismo canto empleando el algoritmo y el modelo de tubos como tracto superior que filtra el sonido emergente.

Comprobación modelo mejorado

En vistas del resultado anterior fuimos a por mejoras en el modelado para recuperar lo que consideramos aquellas características tímbricas que nos faltaron para lograr una síntesis realista. Las principales falencias tímbricas que pudimos observar a partir de la comparación de sonogramas y transformadas de Fourier de los cantos fueron dos.

Por un lado la mayoría de los sonidos del canto del Diamante mandarín presentan 3 bandas enfatizadas en el espectro en la banda de 0 a 10 KHz . Las dos banda ubicuas en todas las sílabas son aquellas cercanas a 2,5 y 7,5 KHz que se recuperan con el modelo de tubo abierto-cerrado. Por lo que desarrollamos el **Modelo de OEC** relatado en la sección anterior, que aporta al sistema la tercer resonancia variable en el tiempo de forma operacional y le da mayor similitud anatómica al modelo.

La otra característica tímbrica faltante fue la “rugosidad” en el sonido, es decir el canto sintético se encontraba falto de imperfecciones. Eso se debe a la falta de ruido en las señales reconstruidas que por construcción son el resultado de una unión suave de puntos (ver Capítulo 2). Por lo tanto la forma de recuperar esa propiedad tímbrica vinculada a una “rugosidad” la modelamos como un ruido en el parámetro $\beta(t)$. La racionalidad de esa inclusión se puede observar en la figura 2.7, donde la medida fisiológica de presión es suave y la de actividad del vS es ruidosa. Eso se debe al proceso de medición donde para obtener datos de actividad muscular se realizan EMGs (electromiografía) y se extrae luego la envolvente de la señal rectificada manteniendo pequeñas fluctuaciones rápidas en la señal [Mindlin et al., 2003, Keynes et al., 2001]. Asimismo la actividad muscular de un músculo pequeño como el vS se puede pensar como ruidosa ya que al contar con pocas fibras las fluctuaciones no se llegan a cancelar estadísticamente.

Con estas dos inclusiones en el modelado para obtener mejoras tímbricas repetimos el experimento. En la figura 3.4 se muestra la nueva síntesis que obtuvimos del mismo canto. Se puede notar la similitud entre los sonogramas, en la secuencia silábica y las frecuencias fundamentales, sin embargo no resulta en este esquema de medidas diferente a la síntesis anterior.

Repetimos el experimento de probar el modelo mediante neuronas selectivas de HVc con síntesis generadas con el modelo mejorado. El resultado se observa en la figura 3.5, los 4 paneles representan cada uno de los estímulos, las barras debajo de cada rótulo representa la duración del estímulo. Los paneles de la derecha que representan al conespecífico (CON) y al reverso (REV) no muestran ningún tipo de actividad, mientras que a la izquierda el canto propio (BOS) si, por lo que se trata de una neurona de HVc que presenta selectividad. Asimismo los dos paneles de la izquierda, que representan al canto propio (BOS) y al modelado (mBOS),

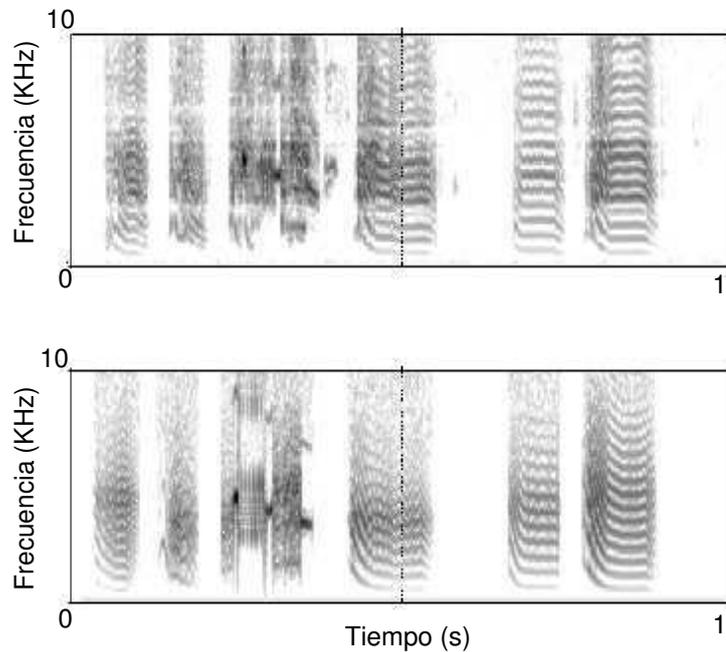


Figura 3.4: La figura muestra dos sonogramas, el superior corresponde a la grabación de un canto real de Diamante mandarín, el inferior corresponde al canto obtenido sintéticamente del mismo canto empleando el algoritmo y el modelo mejorado con la inclusión de la OEC como tracto superior dinámico que filtra el sonido emergente. Asimismo se incorporaron en este síntesis las pequeñas fluctuaciones en el parámetro que representa la actividad muscular.

presentan actividad y estructura temporal de actividad e inhibición similar, por lo que concluimos que esta neurona selectiva identifica al canto sintético como canto propio. En los cuatro paneles se muestra la respuesta al tercer motivo de la presentación del estímulo.

El resultado demuestra que el modelo con las inclusiones del tracto vocal más detallado y el ruido en la actividad de vS es lo suficientemente realista para que el ave en un nivel neuronal lo considere como propio. Lo cual nos permite establecer una jerarquía en términos del modelado que indica que un modelo con una fuente de sonido y un tubo es insuficiente, pero que un modelo con una fuente sonora, un tubo y una caja resonante si. Este resultado es muy relevante al momento de pensar el aporte que puede realizar el modelado físico a sistemas biológicos. La capacidad de establecer jerarquías en un sistema complejo y de muchas variables

3.2 Pertinencia de los modelos en términos de respuesta neuronal

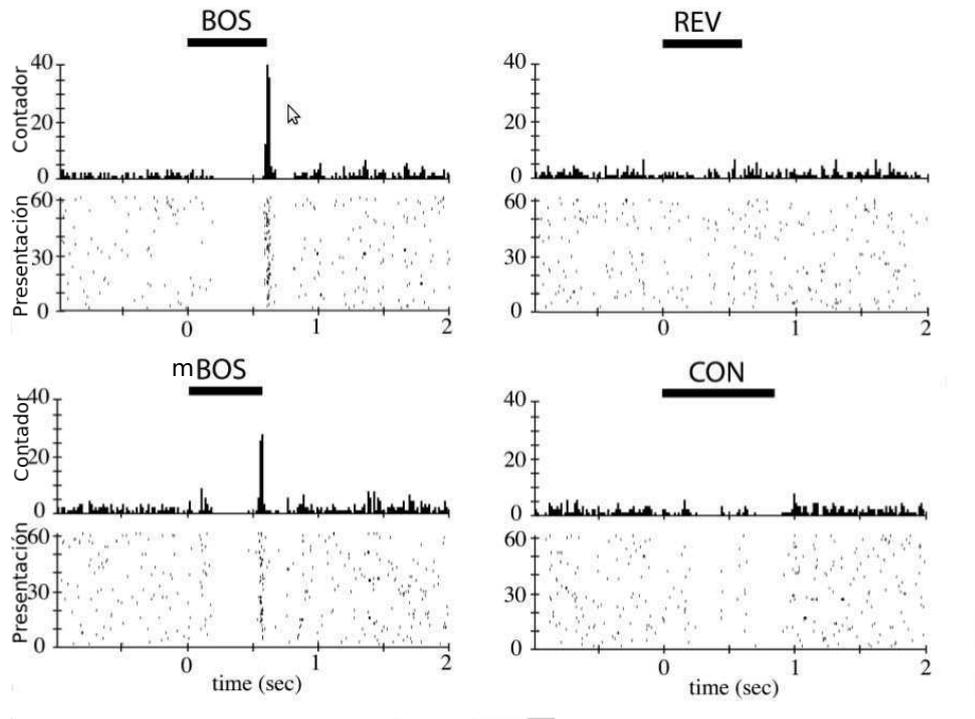


Figura 3.5: Se muestran 4 paneles cada uno de ellos con 2 gráficos: el superior el conteo de de los disparos a lo largo de las presentaciones; el inferior los disparos en función del tiempo en cada presentación. Las barras negras horizontales debajo del nombre del estímulo muestran la duración del mismo y el momento en el que fue presentado. Los paneles de la izquierda muestran la respuesta selectiva al canto propio (BOS) y la respuesta elicitada ante la presentación del canto sintético (mBOS), se observa como presenta la misma estructura temporal de respuesta e inhibiciones. En los paneles de la derecha se muestra la respuesta ante la presentación de los estímulos de canto conespecífico (CON) y canto reverso (REV), se observa la escasa respuesta neuronal a lo largo del tiempo, así como la ausencia de estructuras en dicha respuesta.

resulta determinante a la hora de comprender el sistema y abordar siguiente preguntas sobre el mismo. En este sistema y mediante este tipo de comprobación neuronal hemos encontrado una forma de preguntarle al ave al respecto de estas jerarquías que establecemos desde el modelado.

En [Amador et al., 2013] realizamos de forma sistemática esta comprobación sobre un total de 5 aves y 30 neuronas donde la respuesta promedio obtenida al mBOS fue de $58\% \pm 8\%$ de la respuesta del BOS. En algunos casos ($N = 3$) la

respuesta al mBOS ($valor Z > 0,98 \pm 0,08$) no fue significativamente diferente a la obtenida al BOS (usando un t-test $P > 0,8$). La figura 3.6 muestra un ejemplo de BOS y mBOS con sus sonogramas y actividad neuronal de una neurona proyectora selectiva de HVC para la presentación de los 3 motivos. La selectividad de dicha neurona se comprobó ya que no se obtuvo respuesta al REV ($0,01 \pm 0,01$) ni al CON ($0,05 \pm 0,02$) (t-test $P < 0,001$). Las 30 neuronas mostraron selectividad neuronal y fuerte respuesta al mBOS, una mitad ($N = 15$) respondió con breves ráfagas de disparos neuronales (REV: $0,01 \pm 0,01$; CON: $0,04 \pm 0,02$; mBOS: $0,48 \pm 0,07$; $P < 0,01$) y la otra ($N = 15$) lo hizo con respuesta tónica (REV: $0,01 \pm 0,01$; CON: $0,05 \pm 0,03$; mBOS: $0,65 \pm 0,09$; $P < 0,01$). Los patrones de disparo permitieron a su vez diferenciar a la neuronas proyectoras que responden con ráfagas de disparos (HVC_p), de las interneuronas que lo hacen de forma tónica (HVC_i) [Hahnloser et al., 2002, Kozhevnikov and Fee, 2007].

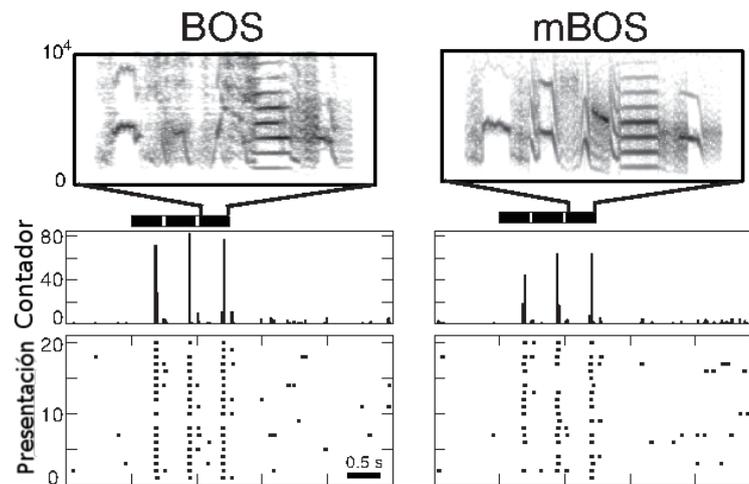


Figura 3.6: Se muestra la respuesta al canto propio (BOS) y al canto sintético modelado (mBOS) con la respuesta a las 3 presentaciones del motivo en ambos casos. En la parte superior se muestra el sonograma del estímulo presentado en cada caso, luego se muestran el conteo de los disparos a lo largo de las presentaciones y los disparos en cada presentación. Se observa como la respuesta en ambos casos es similar en actividad y estructura temporal.

De esta forma hemos probado la pertinencia del modelado del órgano fonador

3.3 Sintonización de respuesta neuronal a variaciones en los parámetros estáticos

aviar con su consecuente filtrado mediante tres puntos de vista distintos. Por un lado en el Capítulo 2, he contado cómo en términos de dos parámetros acústicos (AFF, SCI) el modelo para canto de Diamante mandarín era apropiado para recuperar la relación existente entre ambos. Por el otro en el mismo Capítulo, he contado cómo en términos de los gestos motores el protocolo para sintetizar canto realista con el modelo conduce a reconstrucciones de los mismos con una alta correlación con las medidas experimentales. Finalmente en esta sección he contado cómo en términos de respuesta neuronal el modelado logra elicitación de respuesta comparable con la obtenida por el canto propio. Asimismo pudimos probar qué tipo de modelado era suficientemente adecuado para generar síntesis realista capaz de elicitación de respuesta. Por lo tanto, gracias a estas comprobaciones pudimos establecer la síntesis de canto realista mediante el modelo y su consecuente reconstrucción de gestos motores como una herramienta válida y poderosa para abordar el estudio del sistema de canto.

3.3. Sintonización de respuesta neuronal a variaciones en los parámetros estáticos

En la sección anterior describí en detalle cuál fue el procedimiento empleado para comprobar la validez del modelo de siringe asociado a distintos esquemas de tracto vocal superior. Los resultados obtenidos fueron categóricos, para el esquema de un tubo la respuesta neuronal fue nula y para el esquema de tubo asociado a una cavidad resonante la respuesta fue similar a la de generada por el canto propio. Entonces nos preguntamos cómo se comporta la selectividad neuronal ante variaciones de los parámetros estáticos que llevan la síntesis de canto de un caso al otro. Llamamos parámetros estáticos a aquellos parámetros que no definen la dinámica del canto como ser la secuencia silábica y la frecuencia fundamental (α y β), sino a aquellos que aportan cierta variación tímbrica. Por construcción los dos parámetros estáticos que se presentan como candidatos a estudiar son aquellos que generaron el salto cualitativo de un modelo a otro: el agregado de ruido en $\beta(t)$ y la presencia de la resonancia de la OEC.

A cada una de las 5 aves le presentamos el canto propio (BOS), el canto propio

en reversa (REV), un canto conespecífico (CON) y series de entre 9 y 12 variantes del canto modelado cubriendo una grilla para valores de los parámetros estáticos. Modificando estos parámetros generamos cantos sintéticos con la misma dinámica pero con propiedades tímbricas diferentes. Las variantes fueron creadas de manera sistemática modificando la disipación del resonador de Helmholtz representando la presencia de la OEC y el nivel de ruido adicionado al parámetro reconstruido $\beta(t)$. De esta forma articulando estos dos parámetros podemos barrer un amplio rango de cantos sintéticos. Por ejemplo cuando la disipación de la OEC es grande y el ruido es nulo el filtrado aplicado tiene sólo dos resonancias y el canto no presenta rugosidad, situación análoga al esquema de un tubo. O bien, disminuir mucho la disipación de la OEC y aumentar el ruido da lugar a una resonancia exagerada de la OEC y un canto muy rugoso.

Para encontrar de qué manera sintoniza la selectividad neuronal ante la variación de estos parámetros, grabamos 30 neuronas con respuesta auditiva estadísticamente significativa. Cada una recibió entre 10 y 20 repeticiones de cada estímulo y se comparó con el valor Z obtenido para el BOS. Las neuronas que presentan una fuerte respuesta al canto propio, en la figura 3.7b) se muestra un ejemplo de ello, la selectividad de la misma se comprueba estadísticamente comparando con la nula respuesta al REV y al CON. Gran parte de las variaciones de canto sintético (mBOS) elicitaban respuesta neuronal de diferente magnitud, pero con la misma estructura temporal que la respuesta al BOS, como se muestra en los paneles a) y b) de la figura 3.7. Los ejes indexados de 0 a 10 de a pasos de 5 representan los valores de los parámetros del modelo que dan cuenta del ruido y de la disipación de la OEC. El rango del parámetro del ruido lo variamos de 0 a 40, representando 0 la ausencia de ruido y 5 el valor donde la fluctuación introducida en el parámetro representa el 2,5% del valor de $\beta(t)$. El rango de los valores que toma el parámetro de la disipación va de 0 a 10, donde 0 representa la disipación que hace comparable la resonancia de la OEC a las resonancias del tubo. Aumentar el valor de este parámetro implica aumentar la disipación de la OEC y por ende disminuir la presencia de la tercer resonancia. En algunos casos, el valor de disipación 5 ya era suficiente para dejar de elicitar respuesta.

La mayor respuesta la obtuvimos para mBOS(5,0) lo que significa ruido = 5 y

3.3 Sintonización de respuesta neuronal a variaciones en los parámetros estáticos

disipación de $OEC = 0$, por lo que esos valores fueron elegidos como los indicados para generar síntesis realista. En el panel c) de la figura 3.7 se muestra cómo es la respuesta neuronal sobre el total de las neuronas a medida que cambian los parámetros estáticos. Se observa la capacidad de sintonizar respuesta auditiva selectiva ante variaciones del canto sintético, en particular cómo en el espacio determinado por el ruido y la disipación decae significativamente en ambas direcciones partiendo del $mBOS(5,0)$. Asimismo se observa una diferencia en ambos ejes, ya que la caída en la respuesta es mucho más pronunciada en el eje que crece la disipación respecto del eje que describe la variación del ruido. Esta capacidad de sintonizar la respuesta auditiva de neuronas de HVC a variaciones del BOS es una pregunta que se realizó en trabajos previos [Margoliash, 1983, Margoliash, 1986, Margoliash and Fortune, 1992, Theunissen and Doupe, 1998], pero en este trabajo hemos cambiado el enfoque. Propusimos estudiar la capacidad de sintonizar respuesta auditiva generando cantos variados mediante movimientos posibles de parámetros motores, cuando en los trabajos anteriores se buscaban variaciones en las características acústicas de los cantos.

Esta transformación de coordenadas de un sistema multidimensional acústico a un sistema de baja dimensión de coordenadas motoras es una de las propuestas más innovadoras que plantea nuestro trabajo y permite entender el sistema del canto desde otro punto de vista. Desde este enfoque y basándonos en los resultados que validan el modelo, fuimos a entender cómo funciona parte del sistema de canto interpretándolo en términos de coordenadas motoras. Ese trabajo es el que contaré en el siguiente capítulo.

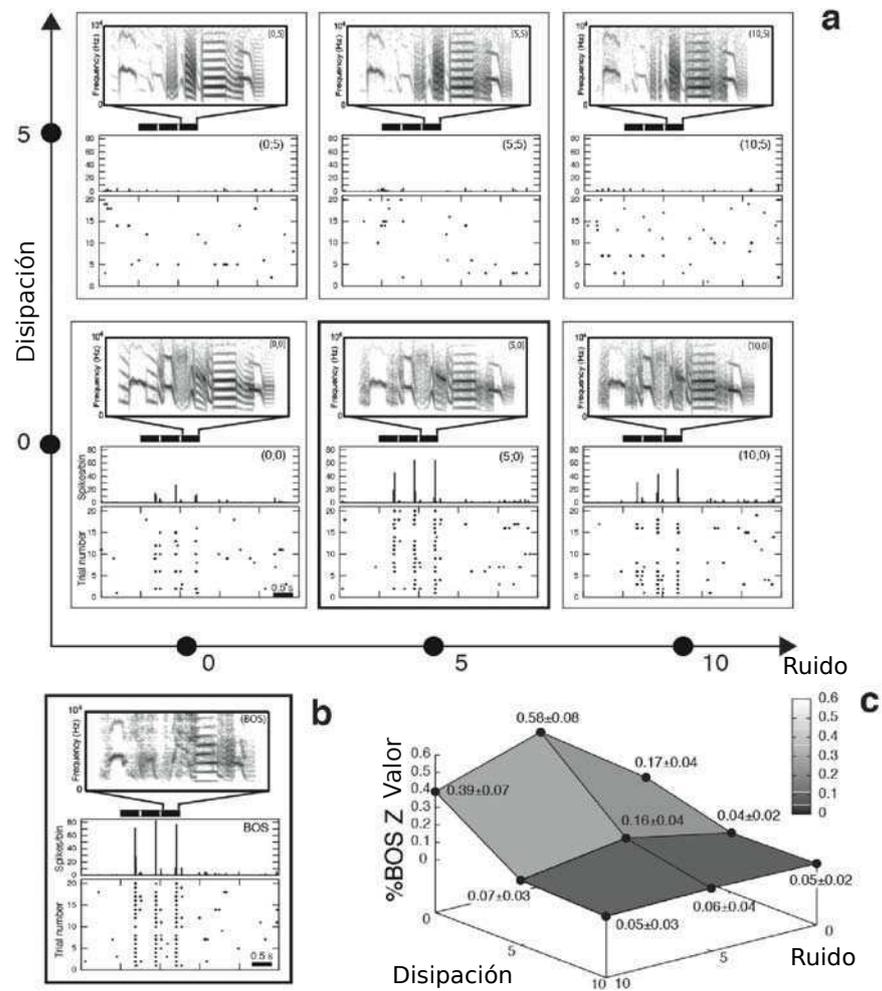


Figura 3.7: a Se muestra una grilla de respuesta neuronal ante la presentación de distintos cantos sintéticos (mBOS) variando dos de sus parámetros estáticos, el ruido en β y la disipación de la OEC. Ambos parámetros definen el espacio de la grilla. b La mejor respuesta ante la variedad de estímulos determinando así los parámetros de la mejor síntesis. Es la misma síntesis mostrada en la figura 3.6. c La superficie que queda definida de respuesta neuronal en el espacio de los parámetros estáticos, se observa como queda determinado así una síntesis óptima y cómo decae a partir de variaciones en los parámetros.

Capítulo 4

Codificación neuronal interpretada en coordenadas motoras.

Los trabajos desarrollados durante una década en el Laboratorio de Sistema Dinámicos, condujeron al desarrollo de un modelo físico de baja dimensión capaz de generar canto sintético realista. Los capítulos anteriores he contado parte de ese desarrollo y cómo se fueron encontrando diversas formas de comprobar la pertinencia de dichos modelados. En particular en uno de los últimos trabajos comprobamos que en términos de respuesta neuronal de cierto núcleo primordial en el sistema del canto (HVC), el modelo logró elicitarse respuesta semejante al canto propio. Por lo tanto se establece el modelado del canto como una herramienta poderosa para ser utilizada en el estudio del canto de aves. Justamente en [Amador et al., 2013] utilizamos la potencia de dicha herramienta para darle una interpretación distinta el sistema del canto de aves que se consideraba en gran parte comprendido y contaba con un paradigma dominante.

4.1. Los gestos motores como unidad del canto

Usando dicho modelado creamos versiones sintéticas de los cantos de las aves a ser estudiadas. Para ello, como he contado en el Capítulo 2, empleamos un pro-

cedimiento algorítmico que reconstruye las señales temporales de los parámetros del modelo en vistas de recuperar las características acústicas del canto. Es importante recordar que el parámetro reconstruido $\alpha(t)$ presenta una fuerte correlación con la presión de los sacos aéreos y el parámetro $\beta(t)$ correlaciona con la actividad del músculo vS. El resultado de este procedimiento para un canto en particular se muestra en la figura 4.1, donde se observan las similitudes en los sonogramas del canto grabado (panel (a)) y del sintético (panel (b)). En el panel (c) se muestran los tramos sencillos dependientes del tiempo que representan a la presión y la tensión reconstruidos a partir del procedimiento de síntesis del canto. Estas dos funciones temporales son las que luego alimentan el modelo no-lineal de los labios que produce un amplio rango de características acústicas propias del Diamante mandarín. Asimismo en el panel (d) se muestra cómo se disponen estos trazos reconstruidos no en función del tiempo, sino ubicados en el espacio de parámetros (α, β) . Vale la pena aclarar, que en el espacio de parámetros para mayor comprensión se graficaron las principales bifurcaciones que presenta la forma normal, como se puede notar se encuentra invertido en ambos ejes respecto al descrito en el Capítulo 2 (ver figura 2.4). Eso se debe simplemente a que durante este trabajo por comodidad se introdujo un cambio de signo a los parámetros α y β en las ecuaciones de la forma normal. Se puede notar cómo las trayectorias en el espacio de presión-tensión cercanas a las bifurcaciones generan cambios en las características del sonido generado con pequeñas modificaciones en su trayectoria. Asimismo se puede observar que existen sonidos que se generan mediante cambios en uno solo de los parámetros manteniendo el otro constante.

De esta forma el canto queda descrito en términos de una secuencia de pequeñas trayectorias en el espacio de presión-tensión que definimos como “gestos”. Cada inicio y final de gesto queda determinado por las discontinuidades en las funciones reconstruidas de presión o tensión. Vale la pena notar que los gestos motores incluyen movimientos que no resultan en sonido, como ser patrones de presión que incluyen mini-respiraciones entre sílabas [Hartley and Suthers, 1989], sin embargo en nuestra definición sólo incluimos aquellos gestos que si generen sonido. Con esta definición de gesto sobre un total de 8 cantos modelados obtuvimos un total de 13 ± 4 gestos por motivo (unidad más larga del canto que repite una determinada secuencia de sílabas). La duración de cada gesto (moda

4.2 Las neuronas proyectoras disparan en los extremos de los gestos motores (GTE)

$= 22,5 \pm 2,5 \text{ ms}$, rango $4 - 142 \text{ ms}$) se distribuye de forma no gaussiana, con un 33 % de los mismos durando menos de 30 ms y una larga cola que corresponde a sonidos de lenta variación como aquellos que presentan frecuencia fundamental constante (llamados “harmonic stacks”). La figura 4.1(e) muestra como se encuentran distribuidos los gestos según su duración.

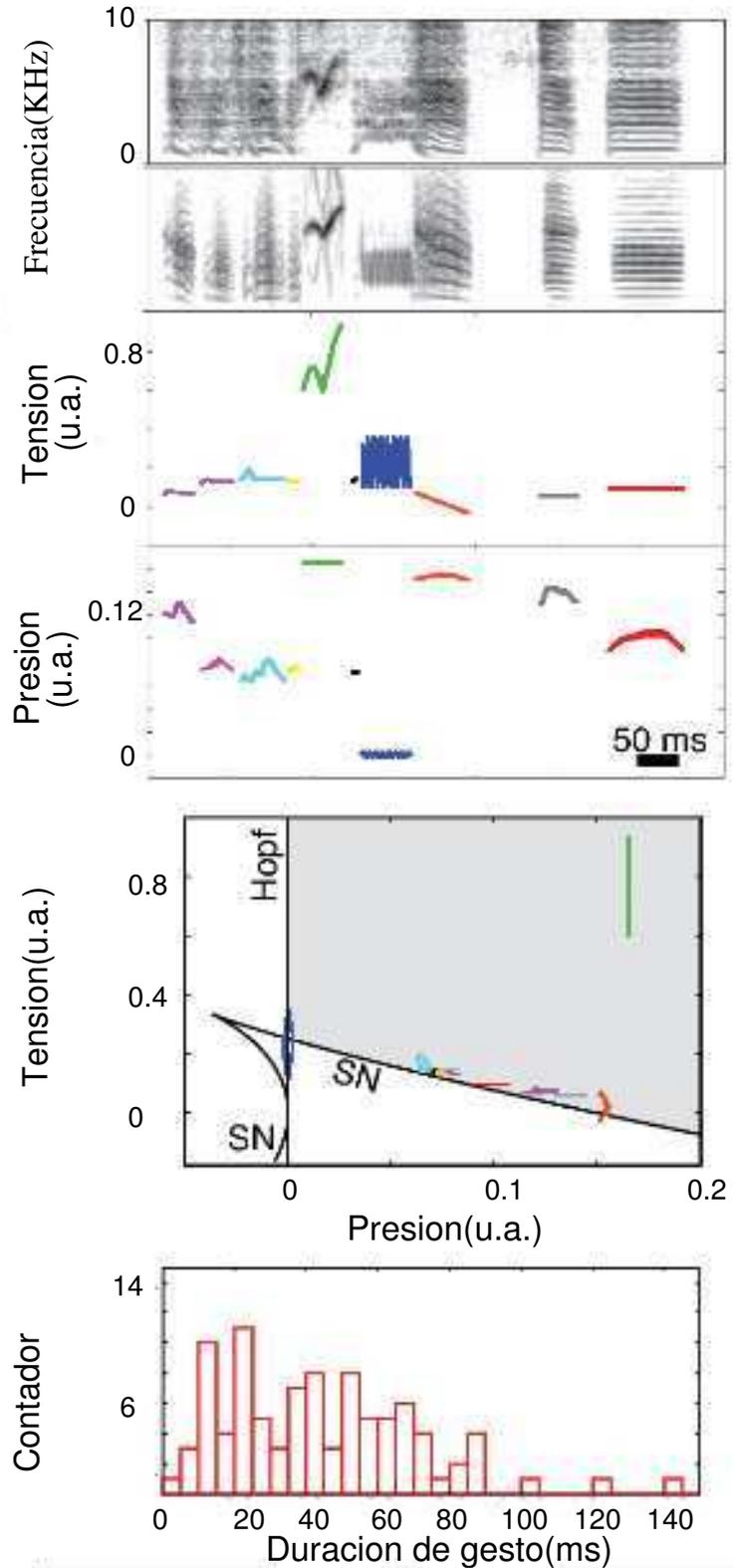
4.2. Las neuronas proyectoras disparan en los extremos de los gestos motores (GTE)

El siguiente paso que dimos en nuestro análisis fue evaluar la actividad de las neuronas proyectoras del HVC en relación a los resultados obtenidos con el modelo dinámico. La obtención de los datos de actividad neuronal de aves dormidas fue realizado por A. Amador empleando el protocolo que fue descrito en el Capítulo 3. Luego comparamos los tiempos en que se producen ráfagas de disparos con las trayectorias de presión-tensión empleadas para sintetizar canto con el modelo (mBOS). De esta forma encontramos una fuerte relación entre los tiempos de disparos de la neuronas proyectoras de HVC (HVC_p) y las señales temporales de los gestos reconstruidos.

4.2.1. Descripción del resultado, aparición de GTE como instancias codificantes

El resultado de analizar esta relación se observa en la figura 4.2 donde se muestran los disparos de dos neuronas (cada una con un color diferente) de un mismo ave en relación con el BOS, el sonograma, la señal temporal del sonido y las señales temporales de presión y tensión reconstruidas. Una de las neuronas (color naranja en la figura) dispara una sola vez durante el motivo y lo hace en un instante particular que es en la transición entre el trino (sucesión de subidas y bajadas de la frecuencia fundamental) y la nota de alta frecuencia constante. La otra neurona representada en verde en la figura dispara dos veces durante el motivo. Una vez lo hace en la transición entre una nota de alta frecuencia y un sonidos de amplio contenido espectral de frecuencia modulada y la otra vez

Figura 4.1: Espectrogramas del canto del ave (a) y el canto sintético modelado (b). El canto es descrito mediante los parámetros reconstruidos $\alpha(t)$ y $\beta(t)$ proporcionales a la presión de los sacos aéreos y la tensión respectivamente. En el panel (c) se muestra esa representación con código de colores para cada fragmento que representa cada sonido. Cada sonido codificado con un color, es generado por trazos continuos en el espacio de parámetros del modelo, cada uno de esos trazos es lo que definimos como un “gesto” (panel (d)). Oscilaciones en la vecindad de la bifurcación de Hopf son sonidos más tonales, mientras que oscilaciones cercanas a la bifurcación de nodo-silla presentan mayor riqueza espectral. Observar que notas espectralmente pobres y de alta frecuencia son muy lejanas a la bifurcación de nodo-silla como es el caso de gesto verde. La región en gris representa la zona del espacio de parámetros donde se producen oscilaciones. En el panel (e) se muestra la distribución de la duración de los gestos para las 5 aves estudiadas.



4.2 Las neuronas proyectoras disparan en los extremos de los gestos motores (GTE)

en el instante en que durante una nota de alta frecuencia la presión alcanza el máximo valor. Este tipo de relación entre los tiempos de disparo neuronales y gestos motores lo encontramos para 14 de las 15 neuronas HVc_p medidas para las 5 aves. El único caso donde no observamos dicha relación fue en una neurona cuyo disparo ocurrió en un intervalo entre sílabas. Interpretamos que la aparición de disparos en momentos de silencio se puede deber a que dicho disparo está asociado a las mini-respiraciones durante el canto. Como he comentado en la sección anterior, sólo modelamos los gestos en los momentos de fonación por lo que sólo consideramos para el análisis aquellos disparos ocurridos en tales momentos. De esta forma se analizaron un total de 17 ráfagas de disparos provenientes de 14 neuronas, que se muestran entre la figura 4.2 y figura A.1.

Analizando sistemáticamente la respuesta de las neuronas de HVc_p en comparación con lo que hemos definido como gesto motor, observamos que preferentemente los disparos ocurren en los extremos de dichos gestos (GTE, por su sigla en inglés). Por definición cada gesto motor tiene al menos dos GTE, un inicio y un fin, y a lo sumo dos GTE más, que sucede si la presión y la tensión tienen un máximo absoluto en tiempos distintos durante el gesto. En casos donde no se presenta un máximo absoluto a lo largo del gesto sino varios máximos locales de igual magnitud no consideramos ningún GTE intermedio (ver gesto motor del trino de figura 4.2 representado en azul). De las 17 ráfagas de disparos, 11 (65 %) estuvieron alineadas con los inicios o finales de gestos, 6 (35 %) estuvieron alineadas con un máximo de presión o tensión.

Para comprender más la envergadura del resultado fue necesario realizar una cuantificación de los GTE por canto, y la naturaleza de esto. Sobre el total de los 5 cantos analizados en esta parte del trabajo extrajimos un total de 165 GTE dando un promedio de 28 ± 4 por canto. Asimismo de los 60 gestos, 20(33,4 %) de ellos presentaron sólo los 2 GTE propios del inicio y fin del gesto; de los restantes 30(50 %) presentaron además un único pico en el gesto de presión dando un total de 3 GTE; los demás gestos, 5(8,3 %) presentaron un único pico en el gesto de tensión y otros 5(8,3 %) un pico en el gesto de tensión y otro en el presión en tiempos distintos dando un total de 4 GTE en esos gestos. En la figura A.1 se muestran las trayectorias de presión y tensión reconstruidas para las 4 aves restantes indi-

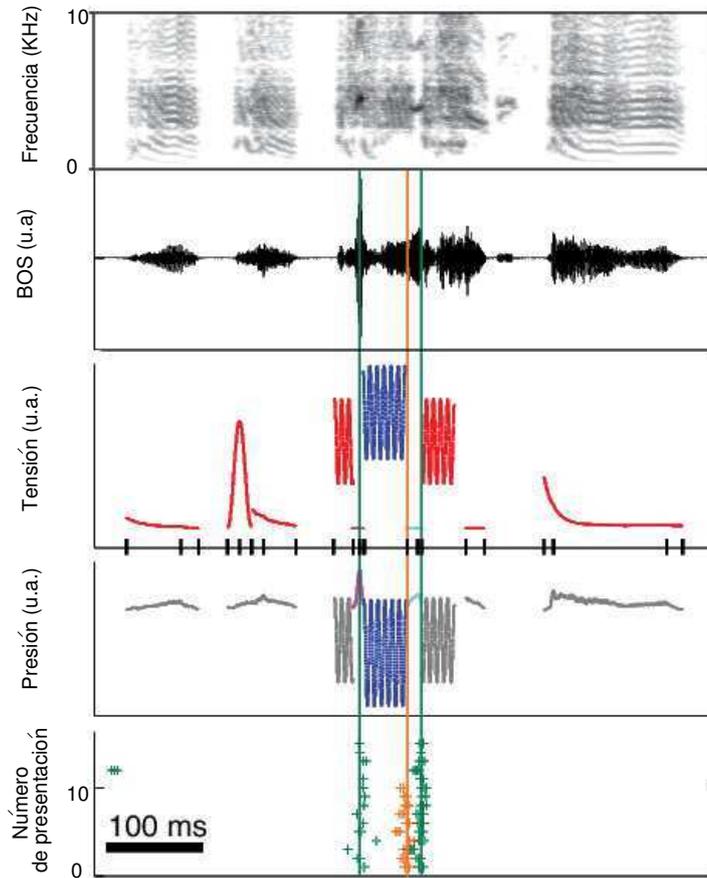


Figura 4.2: De arriba hacia abajo se muestra el sonograma, la señal temporal del sonido del canto grabado, los parámetros reconstruidos de tensión y presión y finalmente los disparos de actividad neuronal de dos neuronas HVC_p en cada presentación del estímulo. Las marcas negras verticales pequeñas corresponden a los tiempos de todos los GTE y las barras verticales de color verde y naranja corresponden al GTE más cercano al instante medio de disparo de cada neurona. Vale aclarar que en este caso una de las dos neuronas dispara dos veces en el motivo por ello existes dos barras verticales verdes.

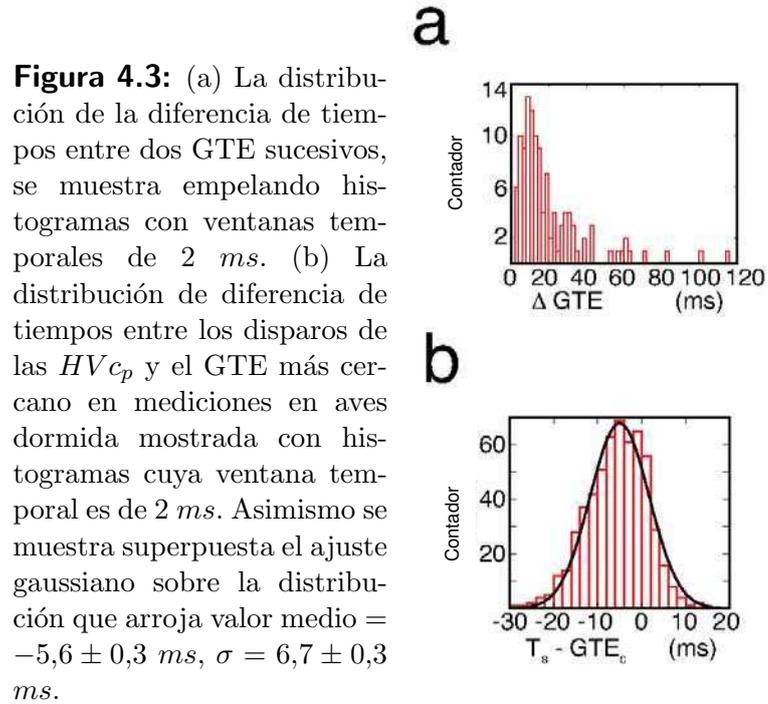
cando con tics verticales en el eje del tiempo la ubicación de los GTE. También se muestra el sonograma del canto propio y del sintetizado con el modelo. Los gestos como se observa corresponden en su mayoría a notas (la unidad de canto más pequeña reconocida por los ornitólogos), pero la actividad motora en los

4.2 Las neuronas proyectoras disparan en los extremos de los gestos motores (GTE)

GTE identificado en los máximos de las trayectorias pueden subdividir notas, como se observa en algunos ejemplos de la figura A.1, o bien en la figura 4.2 donde la ráfaga de disparos de la neurona ocurre en un máximo del gesto de presión en el medio de una nota de frecuencia constante. Este tipo de observaciones resultan indicios significativos para pensar que los patrones de activación de las HV_{c_p} no pueden ser interpretados simplemente en términos de características acústicas.

4.2.2. Estadística sobre los resultados de HV_{c_p}

Para ganar confianza sobre los resultados y analizar la significancia de los mismos realizamos estadísticas sobre los datos. En primer lugar computamos el intervalo temporal existente entre dos GTE sucesivos. La distribución de tiempos es no-gaussiana (moda = 9 ms, rango 4 – 116 ms), con el 66% de los intervalos con duración menor o igual a 30 ms (ver figura 4.3a)). Asimismo computamos la distancia entre cada disparo de cada una de las 17 ráfagas respecto al GTE más próximo. La distribución resultante se observa en la figura 4.3b), allí se muestra cómo se puede aproximar por una distribución gaussiana donde en promedio los disparos neuronales se anticipan al GTE (valor medio = $-5,6 \pm 0,3$ ms, $\sigma = 6,7 \pm 0,3$ ms). Nos preguntamos entonces dada esa distribución cuán significativo era nuestro resultado, para ello generamos conjuntos surrogados de los datos. Cada uno de dichos conjuntos lo construimos adicionando una variable aleatoria a cada una de las neuronas desplazando temporalmente cada uno de los disparos de una neurona en la misma cantidad respetando así los tiempos relativos entre cada uno de ellos. Es decir mantuvimos tanto la estructura interna de la ráfaga de disparos como la varianza de disparo de la neurona física medida. Restringimos la suma del valor aleatorio a aquellos casos donde los disparos corridos no coincidieran con los intervalos entre sílabas, con excepción de dos ventanas de 5 ms antes y después de cada sílaba. En los casos donde algún disparo era empujado hacia un silencio entre sílabas, acomodamos dicho disparo hacia el otro lado de la sílaba. Así obtuvimos poblaciones surrogadas con la misma cantidad de disparos por sílaba, igual estructura interna y cantidad total de disparos (554) igual a la distribución original de los datos. De esta manera construimos 1000 poblaciones



surrogadas cada una con 554 disparos.

Una vez construidas las poblaciones subrogadas estudiamos la significancia del resultado obtenido de los datos. Para ello computamos la distancia entre los cada uno de los disparos de la población subrogada y el GTE más cercano. Vale aclarar que cada uno de los disparos pertenecientes a una misma ráfaga fueron asociados al mismo GTE, que fue elegido como aquel cuya distancia al tiempo medio de la ráfaga era la menor posible. La distribución promedio de las diferencias de tiempo entre los GTE y los disparos surrogados se superpusieron a modo de comparación con la distribución obtenida por los datos y observamos que las colas de la primer distribución eran considerablemente mayores. Dicha superposición se puede observar en uno de los paneles de la figura A.3 ubicada en el apéndice A.

Luego para cada una de las 1000 distribuciones subrogadas de diferencias de tiempos realizamos comparaciones estadísticas con las distribuciones de los datos. Por construcción cada una de las distribuciones comparadas cuentan con la mis-

ma cantidad de elementos y las comparamos mediante un F-test para rechazar la hipótesis nula de que ambas distribuciones son análogas por concordar sus varianzas. Obtuvimos una probabilidad promedio de concordancia en las varianzas de las distribuciones de $P < 0,045$, la figura A.4 muestra en el panel correspondiente cómo va evolucionando el promedio a medida que se computan más F-test entre ambas distribución. La probabilidad P resulta menor que 0,05 con lo que podemos afirmar que ambas distribuciones son significativamente diferentes. En términos de la hipótesis nula que estamos queriendo demostrar este valor indica que la distribución de las diferencias de tiempo de los datos resulta significativamente diferente a distribuciones generadas aleatoriamente. Es decir, el resultado obtenido de que los disparos de HVC_p están alineados con los GTE es estadísticamente distinto al resultado de generar disparos al azar durante el canto, lo que permite ganar confianza sobre los resultados observador en los datos [Amador et al., 2013].

4.3. Las interneuronas son inhibidas en los GTEs

De igual manera que procedimos con las neuronas proyectoras (HVC_p) lo hicimos con las interneuronas (HVC_i) para estudiar la existencia de relación entre la actividad neuronal de éstas y los GTE. Las HVC_i presentan mínimos y máximos locales de actividad tónica a lo largo del canto del ave, nosotros observamos relación sólo entre los mínimos locales y los tiempos de los GTE.

4.3.1. Procesamiento de los datos de actividad tónica

Para caracterizar la relación que identificamos entre los mínimos de la actividad y los tiempos de los GTE fue necesario reconstruir la actividad tónica de una neurona a lo largo del canto. Esto se debe a que la intensidad de la respuesta tónica varía a través de los motivos presentados [Sutter and Margoliash, 1994], por lo que para evitar falsos mínimos fue necesario desarrollar un protocolo para obtener dicha actividad de forma confiable. Este protocolo consistió en tomar de cada estímulo acústico presentado el segundo o tercer motivo cuya respuesta es más intensa y luego para cada uno de ellos realizar una segmentación de los

disparos en ventanas de 10 *ms*. A su vez el protocolo de medición neuronal, como he contado en el Capítulo 3, incluye la presentación de cada estímulo entre 20 y 25 veces, por lo que para cada neurona computamos un promedio de todas las presentaciones. Luego la respuesta promedio con 1 *ms* de resolución fue suavizada empleando un filtro de Savitsky-Golay (regresión polinómica local) [Press, 2007] usando ventanas de 21 puntos superpuestas. De esta forma recuperamos la actividad de las HV_{c_i} a lo largo del canto de forma confiable y repetible para todas las neuronas del experimento y de esta señal suavizada computamos los mínimos. La figura 4.4(a) muestra el sonograma de un canto, los gestos de presión y tensión reconstruidos y la respuesta neuronal computada como relaté más arriba con sus mínimos indicados. Las barras verticales indican los instantes donde están los GTE que coinciden con algún mínimo de la actividad de esta HV_{c_i} . Se puede observar que no todos los GTE coinciden con un mínimo en la actividad de HV_{c_i} , pero a lo largo de las $n = 15$ HV_{c_i} identificamos una fuerte alineación entre los tiempos de los mínimos y los tiempos de los GTE.

4.3.2. Estadística sobre los resultados de HV_{c_i}

Para cuantificar cuán significativa resulta la alineación de los mínimos de actividad con los GTE realizamos procedimientos estadísticos similares a los realizados par el caso de HV_{c_p} . En primer lugar computamos la diferencia de tiempos existente entre el mínimo de actividad ocurrido durante los momentos de fonación y el GTE más próximo. La distribución de dichos tiempos es aproximadamente gaussiana con valor medio = $-0,82 \pm 0,6$ *ms* y $\sigma = 7,3 \pm 1,4$ *ms* y se muestra en la figura 4.4(b). Comparamos esta distribución obtenida de los datos con distribuciones obtenidas de mínimos posicionados de manera aleatoria a lo largo del canto.

Al igual que para las HV_{c_p} generamos poblaciones de mínimos de actividad subrogadas, para ello adicionamos una cantidad de tiempo aleatoria a cada mínimo obtenido de los datos, y ajustamos los corrimientos de forma tal de mantener la misma cantidad de mínimo por sílaba para cada canto. De esta forma generamos 1000 poblaciones subrogadas cada una de ellas con la misma cantidad de mínimos que la población de mínimos extraída de los datos (152) y com-

4.3 Las interneuronas son inhibidas en los GTEs

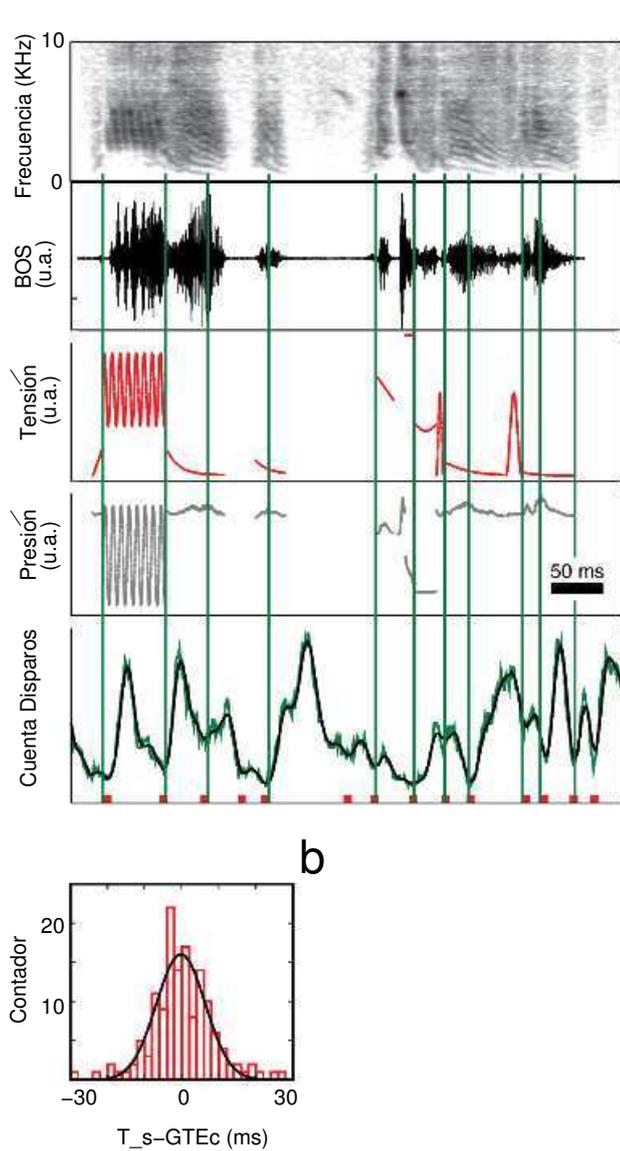


Figura 4.4: (a) De arriba hacia abajo se muestra el sonograma, la señal temporal del sonido del canto grabado, los parámetros reconstruidos de tensión y presión y finalmente el conteo de disparos de actividad neuronal de una HVC_i . Este conteo de disparos se muestra procesado con ventanas temporales superpuestas de 10 ms sobre 20 repeticiones (línea verde) y luego suavizada mediante un filtro de Savitsky-Golay (línea negra). Asimismo los cuadrados rojos representan los tiempos de los mínimos de actividad neuronal de HVC_i de la señal procesada y las barras verticales indican las posiciones de los GTE. (b) La distribución de la diferencia de tiempos entre los mínimos de actividad de HVC_i y el GTE más cercano (GTEc), representada por histogramas con ventanas de tiempo de 2 ms. Superpuesta se muestra el ajuste gaussiano sobre los datos que arroja un valor medio = $-0,82 \pm 0,6$ ms y $\sigma = 7,3 \pm 1,4$ ms.

putamos la diferencia de tiempo de cada mínimo subrogado con el GTE más cercano. Comparamos la distribución subrogada de tiempos promedio obtenida con la distribución de los datos y notamos una diferencia notoria entre las varianzas de ambas (ver el panel correspondiente de figura A.3). Comparamos de forma estadística para cuantificar esta observación cada una de las 1000 distribución subrogada con la de los datos mediante un F-test, en busca de rechazar la hipótesis nula que ambas distribuciones tienen varianzas coincidentes. Obtuvimos de cada comparación una probabilidad P de que coincidan las varianzas que fuimos promediando a lo largo de las 1000 distribuciones subrogadas, el resultado de esta comparación se muestra en la figura A.4 donde se observa que el valor obtenido es $P < 0,016$. Este valor menor a 0,05 muestra que ambas distribuciones son significativamente distintas y por lo tanto poblaciones de mínimos aleatorios generan distribuciones diferentes a las que se obtienen de los datos. Este resultado sugiere que las interneuronas podrían estar ayudando a ajustar el tiempo de disparo de las proyectoras, donde la actividad de las proyectoras inhibe a las interneuronas cuya actividad continua ayuda a dar forma a la siguiente ráfaga de HVC_p .

Luego estudiamos de igual manera la relación entre los máximos de la actividad tónica y los GTE. Para ello sobre la señal de actividad neuronal calculamos los máximos de cada neurona y obtuvimos así una población de máximos considerando las 15 interneuronas medidas de las 5 aves. Luego, al igual que el estudio de los mínimos de actividad, computamos la distribución de la diferencia de tiempos entre cada máximo y el GTE más cercano. En la distribución de mínimos el 67% de los datos están en el rango $\pm\sigma$ y el 90% en uno de $\pm 2\sigma$, sin embargo en la de los máximos el 63% están en el rango $\pm\sigma$ y sólo el 83% en un rango de $\pm 2\sigma$, por lo que se nota la diferencia de varianzas de ambas distribuciones (ver figura A.5). Repitiendo el procedimiento de comparación generamos poblaciones subrogadas aleatorias con igual cantidad de máximos por sílaba y se comprobó la no-significancia de la distribución de los datos respecto a las subrogadas (F-test $P = 0,55$), por lo que no podemos afirmar nada al respecto de la relación entre los máximos y los GTE.

Luego de obtener estos resultados realizando los experimentos en aves dormi-

das, surgió la necesidad de conocer qué sucedía entonces con el ave cantando, por lo que resultó imperante realizar mediciones sobre aves despiertas. En la sección siguiente contaré el desarrollo y los resultados de dichas mediciones.

4.4. La representación de los gestos durante el canto

Como he relatado en secciones anteriores, trabajos previos mostraron el fenómeno de réplica entre el ave cantando y el ave dormida en Diamante mandarín para el núcleo RA. Por otro lado trabajos sobre Diamante bengalí y Gorriónes de pantano, probaron que los patrones de disparo de neuronas de HVc son similares durante la estimulación auditiva y la ejecución del canto. Sin embargo, el resultado para neuronas de HVc en Diamante mandarín aun no fue reportado, por lo que para validar nuestro resultado observado en aves dormidas, se realizaron mediciones neuronales de aves despiertas y cantando.

El protocolo de trabajo fue análogo al realizado para aves dormidas, en el LSD realizamos los modelados de los cantos y en la U. de Chicago A. Amador realizó los experimentos sobre las aves. En este caso la principal diferencia fue que en vez de implantarle en la cabeza un pin sujetador, se le implantó de forma crónica un micromanipulador motorizado que le permitió al ave comportarse libremente. Con el micromanipulador se logró controlar de forma permanente la posición de un electrodo para medir neuronas en el HVc con el ave cantando. Se realizaron mediciones de HVc para $n = 3$ aves cantando, incluyendo un total de 10 neuronas tónicas y 5 neuronas fásicas que dispararon durante el canto (una neurona presentó 2 ráfagas por motivo y se midió para 2 de las 3 aves en experimentación). En la figura 4.5 se muestra un ejemplo de las mediciones obtenidas, en el panel (a), la respuesta en la repetición del motivo; en el panel (b), la extracción de una sílaba sumando su respuesta neuronal en distintos motivos. Estas mediciones confirman que durante el canto la actividad de las neuronas de HVc_p sigue consistiendo en disparos ralos y coincidente con transiciones en los gestos.

Con los gestos reconstituidos mediante el modelado de la canción y la ac-

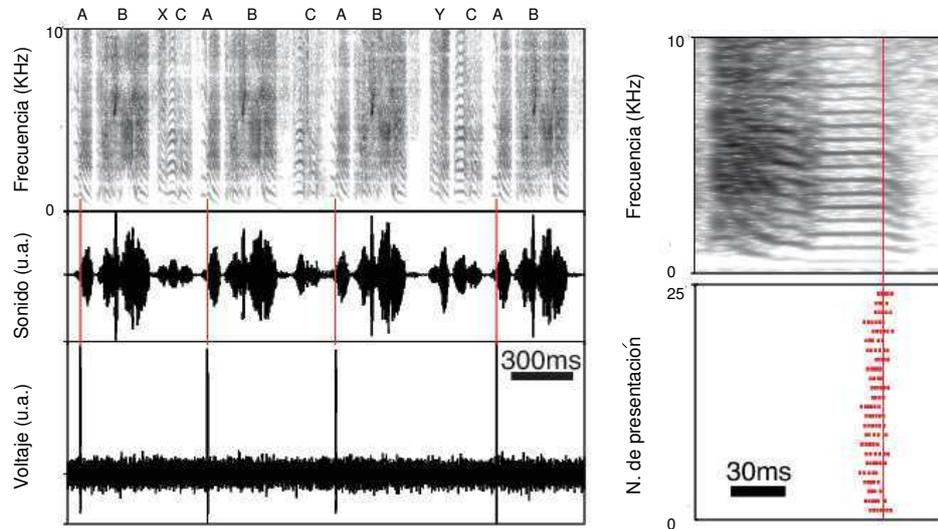


Figura 4.5: **izq** Una neurona HV_{cp} disparando de forma sincrónica con un GTE ubicado en la transición de la sílaba rotulada con la letra A (las letras sobre el sonograma indican distintas sílaba). Se nota como el disparo se sincroniza con el GTE indicado incluso cuando la secuencia silábica e el intervalo de tiempo cambian. **der** Se muestra para un ave distinto las ráfagas de disparo en cada presentación y el sonograma de la sílaba indicando que el disparo sucede en una transición acústica muy sutil que sí se identifica como un GTE.

tividad neuronal medida simultáneamente cuantificamos el resultado obtenido de la sincronización entre GTE y actividad neuronal para aves cantando para las HV_{cp} . Vale aclarar que en este caso, a diferencia de los experimentos en aves dormidas, el modelado del canto se realizó considerando las variaciones observadas en la ejecución del motivo por parte del ave. Computamos la diferencia de tiempo entre la actividad de HV_{cp} y los GTE extraídos, la distribución vuelve a ser aproximadamente gaussiana con valor medio = $-1,35 \pm 0,1$ ms y $\sigma = 4,0 \pm 0,1$ ms como se observa en el panel correspondiente de la figura 4.6, mostrando que la sincronización entre los disparos y los GTE resulta incluso más precisa que en aves dormidas. De igual manera que para la población de disparos obtenidos de las aves dormidas, generamos 1000 poblaciones subrogadas de igual cantidad de disparos (293) y estructura interna que la población de los datos con la finalidad de realizar comparaciones estadísticas. La distribución promedio de la diferen-

4.5 Implicancias de la codificación motora.

cia de tiempo entre el GTE más cercano y los disparos surrogados superpuesta con la distribución original de los datos vuelve a mostrar diferencias en las colas de sus distribuciones (se muestra en el panel correspondiente de la figura A.3). Luego para cuantificar esta diferencia realizamos un F-test entre cada una de las 1000 distribuciones subrogadas y la distribución de los datos para comparar las varianzas dando una probabilidad promedio $P < 0,025$ (ver figura A.4). De esta forma demostramos que ambas distribuciones son significativamente distintas.

Luego el analizamos la relación entre los mínimos de actividad de HVc_i en aves cantando con los tiempos de los GTE extraídos de modelizar los cantos. Aplicando el mismo procedimiento que para el caso de mediciones en aves dormidas obtuvimos la actividad tónica y computamos los mínimos durante el canto. Éstos mostraron nuevamente una sincronización precisa respecto a las instancias determinadas por los GTE. Construimos la distribución de la diferencia de tiempos entre cada mínimo y el GTE más cercano (se muestra el panel correspondiente de la figura 4.6) cuyo ajuste gaussiano sobre la misma dio valor medio = $-0,12 \pm 0,40$ ms y $\sigma = 4,0 \pm 0,4$ ms. Asimismo generamos poblaciones subrogadas de mínimos respetando la cantidad de mínimos por sílaba de cada canto y la cantidad total (95) del conjunto de datos. Luego para cada una de las poblaciones subrogadas computamos la diferencia de tiempos con los GTE más cercanos y obtuvimos distribuciones subrogadas (1000) que comparamos con la distribución de los datos (ver figura A.3). Para cuantificar cuán distintas estadísticamente resultan dichas distribuciones con la distribución de los datos, nuevamente computamos un F-test para cada una de las 1000, dando un valor promedio del $P < 0,002$ (ver figura A.4). Así demostramos que la sincronización entre los mínimos de actividad de las HVc_i y los GTE es significativamente distinto a lo obtenida con poblaciones generadas al azar.

4.5. Implicancias de la codificación motora.

Como he contado en la introducción de este trabajo, anteriormente a estos resultados se pensaba que el tiempo de la actividad neuronal en HVc_p no estaba relacionado con los tiempos propios del canto [Hahnloser et al., 2002, Fiete et al.,

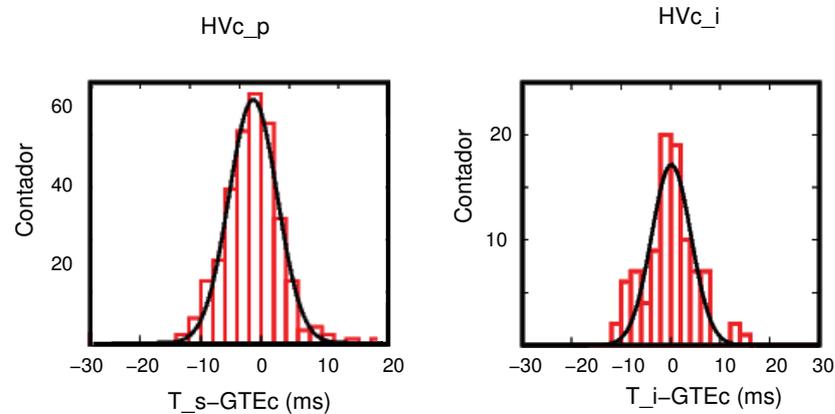


Figura 4.6: Se muestran las distribuciones de las diferencias de tiempos entre: **izq.** los disparos de las neuronas HVc_p y el GTE más cercano (GTEc); **der** los mínimos de actividad neuronal de HVc_i y el GTEc. Cada distribución está construida mediante histogramas con conteos de a ventanas de 2 ms . A su vez se superpone sobre cada distribución el ajuste gaussiano de los datos obteniéndose para la distribución de las HVc_p valor medio = $-1,35 \pm 0,1\text{ ms}$ y $\sigma = 4,0 \pm 0,1\text{ ms}$ y para HVc_i valor medio = $-0,12 \pm 0,40\text{ ms}$ y $\sigma = 4,0 \pm 0,4\text{ ms}$

2004]. La hipótesis imperante en la comunidad de sistema motor, sostenía que la actividad del HCv se produce en ráfaga de actividad rala pero que cada una de estas corresponde a un “tick” de un reloj generador de pulsos uniformemente equiespaciados en intervalos de aproximadamente 10 ms [Fee et al., 2004]. Sin embargo, nuestros resultados muestran que la actividad de HVc_p y la modulación de HVc_i están significativamente sincronizadas con algunos tiempos particulares determinados por los gestos motores. Si bien la secuencia de disparos a lo largo de la población de HVc_p se encuentra temporalmente ordenada como sugieren Fee y colaboradores [Hahnloser et al., 2002], no estaría determinada por una representación temporal en el HVC, sino por los tiempos marcados por la mecánica de la siringe y el tracto vocal. Como he mostrado a lo largo de este capítulo, los GTE presentan una amplia distribución temporal entre ellos, por lo tanto si la actividad de HVC está sincronizada con los GTE, como nuestro modelo muestra, resulta inconsistente con la existencia de una red de neuronas actuando como una cadena de disparos que resultan en actividad equiespaciada uniformemente. La diferencia entre ambos modelos respecto al funcionamiento del HVC tiene además

implicancias en la comprensión de la organización funcional en el sistema del canto, en el mecanismo de aprendizaje y la codificación motora

Una de estas implicancias se basa en considerar que por un lado los gestos tienen gran variación en sus duraciones y estructuras pero el RA sólo recibe información temporal de los instantes de los GTE. Por lo tanto, según este modelo, las componentes que se encuentran “río abajo” en la vía motora (el RA y el tronco cerebral) deberían ser las responsables de generar la información dinámica necesaria para generar la estructura específica dentro de cada gesto. En esa dirección experimentos previos [Williams and Vicario, 1993, Vu et al., 1994] habían mostrado la importancia de la información existente en HVc para la codificación de largas unidades del canto. Por otro lado, los gestos son aprendidos y esto es consistente con las propiedades fisiológicas de las neuronas del HVc, integran sobre cientos de *ms* y sobre múltiples sílabas, realizan sumas no-lineal sobre las sílabas en una secuencia de respuesta pre excitatoria y presentan respuesta selectiva al BOS [Dave and Margoliash, 2000, Margoliash, 1983, Margoliash, 1986, Margoliash and Fortune, 1992, Prather et al., 2009]. Por lo tanto la información en grupos de gestos, como por ejemplo sílabas, puede ser acumulada en este tipo de señales integradas. Otra implicancia vinculada al aprendizaje tiene que ver con que el HVc proyecta hacia el AFP que contribuye al aprendizaje generando variaciones en la actividad de RA que finalmente se traducen en modificaciones en el canto resultante. Esta observación sugiere que las variaciones no son estructuradas en el marco sonoro-auditivo sino en torno a características de los gestos motores.

Una consecuencia fuerte de los resultados obtenidos se relaciona con el rol que cumple el HVc en el sistema del canto. El modelo pre-existente colocaba al HVc en la cima de la cadena de núcleos premotores de donde se comienzan a generar las instrucciones que luego “bajan” por los sucesivos núcleos hasta obtener como resultado un canto. Sin embargo, nuestros resultados muestran que la actividad de HVc está sincronizada, con una pequeña diferencia de tiempo, con los gestos motores que suceden en la periferia, por lo tanto no es consecuente con el rol pre-motor. Esta observación sugiere que las neuronas proyectoras de HVc representan una predicción del comportamiento emergente en ese instante de tiempo, representando un modelo de predicciones capaz de resolver el problema planteado de los retardos en el control sensorio-motor. Actualmente distintas investigaciones

Cap.4 Codificación neuronal interpretada en coordenadas motoras.

están analizando conjeturas sobre cómo se puede articular el rol predictivo del HVc integrado con el resto del sistema del canto, las vías auditivas con su re-alimentación auditiva, el AFP. En particular en el LSD se están realizando un modelado que contempla el carácter predictivo del HVc integrándolo con mediciones realizadas en el laboratorio de gestos motores medidos al someter al HVc a diferentes temperaturas.

Capítulo 5

Automatización del algoritmo

Una vez desarrollado el algoritmo para reconstruir los gestos y generar canto sintético con su pertinencia comprobada en términos de actividad neuronal, vimos la necesidad de desarrollar una automatización del procedimiento. El algoritmo descrito para sintetizar canto es preciso y completamente repetible por cualquier investigador o estudiante tal como está en [Perl et al., 2011, Amador et al., 2013], sin embargo requiere de un entrenamiento por parte del individuo. El desarrollo de programas que realicen toda la tarea de forma automática persiguen el objetivo de brindar a la comunidad de canto de aves una herramienta, sin requerir entrenamiento previo, que permita implementar la representación motora para abordar preguntas sobre el sistema del canto. Por otro lado acelera significativamente los tiempos empleados en la reconstrucción y síntesis permitiendo procesar grandes cantidades de datos y así abordar otro tipo de preguntas que requieran análisis estadístico sobre los resultados.

El algoritmo de reconstrucción de gestos y síntesis consta de tres instancias principales sucesivas cuyo resultado de una es utilizado por la siguiente. La primer etapa consiste en, sobre las grabaciones de canto, obtener la segmentación silábica del mismo y la frecuencia fundamental. La segunda etapa consiste en la reconstrucción de gestos ajustando mediante el modelo dinámico de órgano fonador los valores de los parámetros $\alpha(t)$ y $\beta(t)$ tal que recuperen los valores de frecuencia fundamental. La tercer etapa es aquella en donde se sintetiza el canto mediante el modelo usando las señales temporales de los parámetros reconstruidos. Para la

automatización del algoritmo implementé numéricamente de forma separada las tres etapas en programas computacionales empleando lenguaje de programación de código abierto (C y python).

5.1. Obtención de la frecuencia fundamental

Esta primera etapa del algoritmo de reconstrucción y síntesis representa la más crítica al momento de diseñar un algoritmo automático debido a que busca recuperar información a partir de señales reales grabadas. En particular el objetivo es obtener automáticamente el valor de la frecuencia fundamental como función del tiempo a lo largo del canto, y esto en sí es una tarea compleja debido a que esta varía rápidamente con el tiempo. Esto se debe a que el análisis frecuencia-tiempo se encuentra limitado por el principio de incertidumbre, acotando la posibilidad de computar frecuencias precisas a señales temporales cortas. Asimismo al tratarse de sonidos reales y ricos acústicamente la identificación de una frecuencia fundamental en ese contexto acústico se vuelve particularmente difícil. Para resolver dichas dificultades me basé en los trabajos de T. Gardner y M. Magnasco quienes desarrollaron un método capaz de identificar frecuencias de forma precisa definiendo una frecuencia y un tiempo instantáneos [Gardner and Magnasco, 2006, Lim et al., 2012].

En sus trabajos T. Gardner y co-autores se preguntaron justamente cómo procesa el cerebro la información auditiva considerando que preserva de forma muy precisa la información temporal de las características del sonido a pesar de los rápidos cambios que los sonidos reales presentan. Con esa idea plantearon un método de procesamiento de la información sonora que puede ayudar a comprender la agudeza de la percepción auditiva. En primer lugar sugieren que la cóclea actúa como un segmentador de la señal sonora que recibe, $\chi(t)$, en muchas componentes de la forma $\chi(t, \omega)$ etiquetadas según ω :

$$\chi(t, \omega) = \int e^{-(t-t')^2/2\sigma^2} e^{j\omega(t-t')} \chi(t') dt'. \quad (5.1)$$

donde $\chi(t, \omega)$ se conoce como la transformada de Gabor o la STFT (por su sigla en inglés Short-time Fourier Transform). El parámetro σ es un parámetro tem-

poral que da cuenta de la escala de tiempo de la transformada y es la inversa de la resolución en frecuencia. En coordenadas polares la transformada se puede escribir como $\chi(t, \omega) = |\chi|e^{i\phi(t, \omega)}$, donde el $|\chi|^2$ como una función de (t, ω) es lo que se conoce como sonograma. La fase de la STFT es lo que se encuentra en el exponente de la expresión en polares, $\phi(t, \omega)$, a partir de esta variable generan el cambio de coordenadas que les permite transformar el espacio determinado por (t, ω) al nuevo espacio determinado por las variables instantáneas (t_{ins}, ω_{ins}) . De esta forma con un algoritmo de reconstrucción posterior obtiene una representación en el espacio (t, ω) que llama “contornos“ que resulta más precisa que el sonograma tradicional. Sin embargo, para los fines de nuestro trabajo no avancé sobre todo el método sino que empleé la STFT para aproximarme a la obtención de la frecuencia fundamental. considerando sí, los valores de los parámetros que optimizan el método para canto de Diamantes mandarín según [Lim et al., 2012].

El siguiente paso fue realizar un programa computacional que implemente el cómputo de la STFT para obtener una función del tiempo y la frecuencia ($\chi(t, \omega)$) a partir de una señal temporal obtenida del canto grabado ($\chi(t)$). Para ello es necesario seleccionar ventanas temporales de la señal sonora grabada a una determinada frecuencia de muestreo así como la selección del valor del parámetro temporal σ que representan parámetros determinantes en el cómputo la STFT. Tomando los trabajos realizados T.Gardner, quien ajustó dichos parámetros para optimizar su método al canto de Diamante mandarín, determiné el valor de dichos parámetros. La frecuencia de muestreo optimizada fue de $44,1 \text{ KHz}$, que permite tomar ventanas óptimas de 1024 puntos en torno al instante t en que se computa la STFT. Notar que esta relación entre frecuencia de muestreo y cantidad de muestras por ventana optimizada de forma cuantitativa por Lim y co-autores da por resultado ventanas de aproximadamente 20 ms que son de igual duración que aquellas empleadas por el algoritmo descrito en [Perl et al., 2011]. El parámetro temporal optimizado para realizar un buen cómputo de la STFT en canto de Diamante mandarín es $\sigma = 5 \text{ ms}$. Por lo tanto, diseñé un programa que obtuviera la señal grabada de un canto a $44,1 \text{ kHz}$ para que instante a instante (muestra a muestra del canto digitalizado) tome 512 puntos hacia cada lado y compute la STFT con un filtro de contante $\sigma = 5 \text{ ms}$. De esta forma para cada instante t dentro del canto (incluyendo sílabas e intersílabas) el programa reconstruye de

la STFT la amplitud de cada frecuencia presente en el sonido, obteniendo así un sonograma.

Ahora bien, resta identificar cuál es la frecuencia fundamental en cada instante y determinar a lo largo del canto los momentos de fonación y los de silencio. Para ello realicé un sistema de filtros aplicados sobre el módulo de las STFT instante a instante tanto en amplitud como en frecuencia. Los parámetros de dichos filtros se ajustaron empíricamente a partir de computar sucesivamente la STFT para distintos cantos. Por un lado el filtrado en amplitud funciona a modo de umbral y permite identificar para cada instante si el contenido espectral computado corresponde a momentos de silencio o de fonación, así como impedir la detección de frecuencias fundamentales ficticias. El filtrado en frecuencia, acota la banda de posibles frecuencias fundamentales a aquellas que tienen sentido para el canto de Diamante mandarín. De esta forma se obtiene para cada canto grabado una serie temporal que representa la frecuencia fundamental con igual resolución temporal que el canto. Sobre esta señal se realiza un promedio temporal de a ventanas de entre 100 y 350 puntos para ambos lados del instante t para lograr un trazo suave. En la figura 5.1 se muestran dos ejemplos de la frecuencia fundamental reconstruida y promediada superpuesta a un sonograma realizado con el algoritmo de la librería *Matplotlib* de python. Se observa como la serie temporal de la frecuencia fundamental recupera muy bien el valor de la misma a lo largo del canto y como resuelve el problema de la segmentación silábica. En el apéndice B de esta tesis se encuentra el código en C que realiza esta operación.

Vale aclarar que esta automatización funciona correctamente para la mayoría de los cantos de Diamante mandarín, sin embargo en algunas ocasiones no ajusta bien la frecuencia fundamental. Eso se debe a que en oportunidades los cantos presentan sílabas muy ruidosas o modulaciones muy rápidas de frecuencia que no son bien computadas por este código.

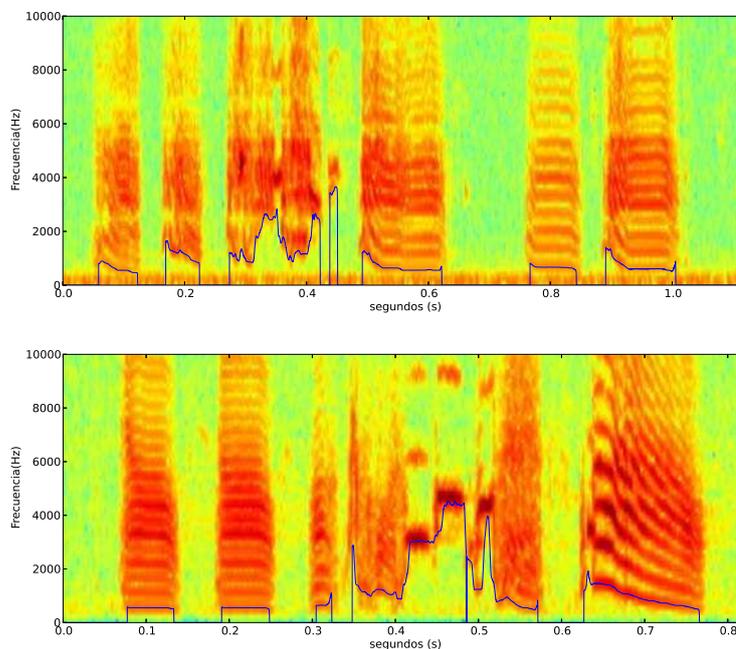


Figura 5.1: La figura muestra dos ejemplos de frecuencia fundamental reconstruida mediante el programa que implementa el cálculo de STFT y una serie de filtros y luego promediada. En azul se marca en ambos paneles las correspondientes frecuencias fundamentales y en escala de colores se muestran los sonogramas correspondientes a cada canto graficados con *Matplotlib*. Notar como el trazo azul se pega muy bien a la frecuencia fundamental del sonograma.

5.2. Reconstrucción de los gestos y síntesis de canto automática

En este caso la señal temporal de las frecuencias fundamentales promediadas pasa a ser la representación del canto una vez realizado el cómputo de la primer etapa. Por lo tanto el objetivo fue crear un programa que mediante el modelo dinámico del órgano fonador aviar reconstruyera los valores de los parámetros fisiológico que controlan al mismo ($\alpha(t)$ y $\beta(t)$). Esta etapa ya había sido programada anteriormente y era usada de forma manual en el algoritmo descrito en [Perl et al., 2011], por lo que simplemente tuve que adaptar el código a nuestra nueva representación del canto en términos de una serie temporal de frecuencia

fundamental.

La experiencia de realizar muchas síntesis y reconstrucción de gestos motores empleando el algoritmo nos permitió realizar observaciones empíricas que fueron implementadas de forma práctica en la automatización. La principal observación consiste en que para obtener un buen ajuste de contenido armónico según la frecuencia fundamental (cuantificado por el SCI) en general es suficiente con fijar el valor de α en $\alpha = -0,15 \pm \epsilon$ con ϵ pequeñas perturbaciones determinadas por las variaciones en la amplitud de la señal sonora. Luego la frecuencia ajustada simplemente con movimientos en β resulta en el contenido espectral y la frecuencia fundamental adecuados. Asimismo, la señal temporal resultante de α como $\alpha(t) = -0,15 \pm \epsilon(t)$ resulta una buena reconstrucción del gesto de presión ya que sigue las variaciones de la amplitud de la señal sonora que correlacionan fuertemente con la señal de presión de los sacos aéreos.

La etapa del algoritmo automatizado para reconstruir los gestos motores consiste entonces en a cada instante t_0 de la señal temporal de frecuencias fundamentales ajustar el valor de β dado el valor de $\alpha(t_0) = -0,15 \pm \epsilon(t_0)$. Para ello utiliza un banco de datos de simulaciones numéricas realizadas con el modelo para una grilla de 100×100 en el espacio (α, β) barriendo un rango de frecuencias tal que cubra las posibles frecuencias fundamentales. Así determina para ese instante t_0 el valor de α y β , que repetido para todo instante del canto termina con reconstruir ambos gestos para todo t con un muestreo igual al de la grabación de canto. Vale aclarar que cuando el programa encuentra un espacio intersilábico, es decir sin sonido, los gestos se mueven hacia la región de no fonación del espacio de parámetros. El resultado de la automatización de la reconstrucción de gestos se puede ver en la figura 5.2 en los paneles inferiores para un ejemplo de un canto cuyo sonograma y síntesis se observan en los paneles superiores. El código correspondiente para esta etapa se encuentra disponible en el apéndice B.

La síntesis de canto se realiza de igual manera que en algoritmo descrito en [Perl et al., 2011], cuyo código se vino desarrollando durante años en el LSD. Éste integra las ecuaciones diferenciales del modelo alimentadas por valores de los parámetros α y β , por lo que tuve que realizar una adaptación del mismo que consistió en que los valores que incorpore para integrar sean los reconstruidos en la etapa anterior. De esta manera se obtiene una síntesis de canto generada por

los gestos motores reconstruidos, que luego es filtrada tal y como fue descrito en el capítulo 3 obteniendo canto tímbricamente realista. La figura 5.2 muestra en los paneles superiores los sonogramas del canto grabado y la síntesis correspondiente.

La calidad de la síntesis depende de lo precisas que hayan resultado las etapas anteriores, por lo que como he dicho antes, esta automatización resulta pertinente para gran parte de los cantos de Diamante mandarín pero no para todos. Por lo tanto surge la necesidad de poner algún criterio que mida cuán pertinente resulta la síntesis automática y si efectivamente puede ser instalada como una herramienta de análisis en coordenadas motoras. Nuevamente se impone como criterio el establecido para la síntesis manual, la respuesta en HVc de neuronas selectivas al canto propio.

5.3. Selectividad en HVc

A partir de la repatriación de A. Amador incorporada como investigadora al Laboratorio de Sistemas Dinámicos, se comenzaron a realizar experimentos de electrofisiología en aves en el país. Para ello, no sólo se incorporó la experiencia y conocimientos adquiridos por ella en su post-doctorado, sino que además se compraron y diseñaron equipamientos adecuados para realizar los experimentos de forma análoga a la realizada en la U. de Chicago. De esta manera fue posible abordar la pregunta sobre cuán pertinente es la síntesis de canto realizada automáticamente en términos de respuesta neuronal en el Laboratorio a la par que consolidan este tipo de procedimientos en el país.

Enmarcado en el contexto de la tesis de licenciatura del Lic. Santiago Boari, bajo la dirección de A. Amador, se realizaron los primeros experimentos de neurofisiología en el Laboratorio midiendo neuronas selectivas en HVc. Para ello se estableció como protocolo el mismo usado en los experimentos realizados en Chicago en aves dormidas. En primer lugar se implanta un pin sujetador en la región posterior de la cabeza del ave que luego es montado en una estructura de soporte que permite fijar al ave mediante el pin (ambas piezas desarrolladas en el departamento de física con diseño propio y realización del tornero). Luego sobre dos posicionadores micrométricos perpendiculares se ubica dicha estructura de fi-

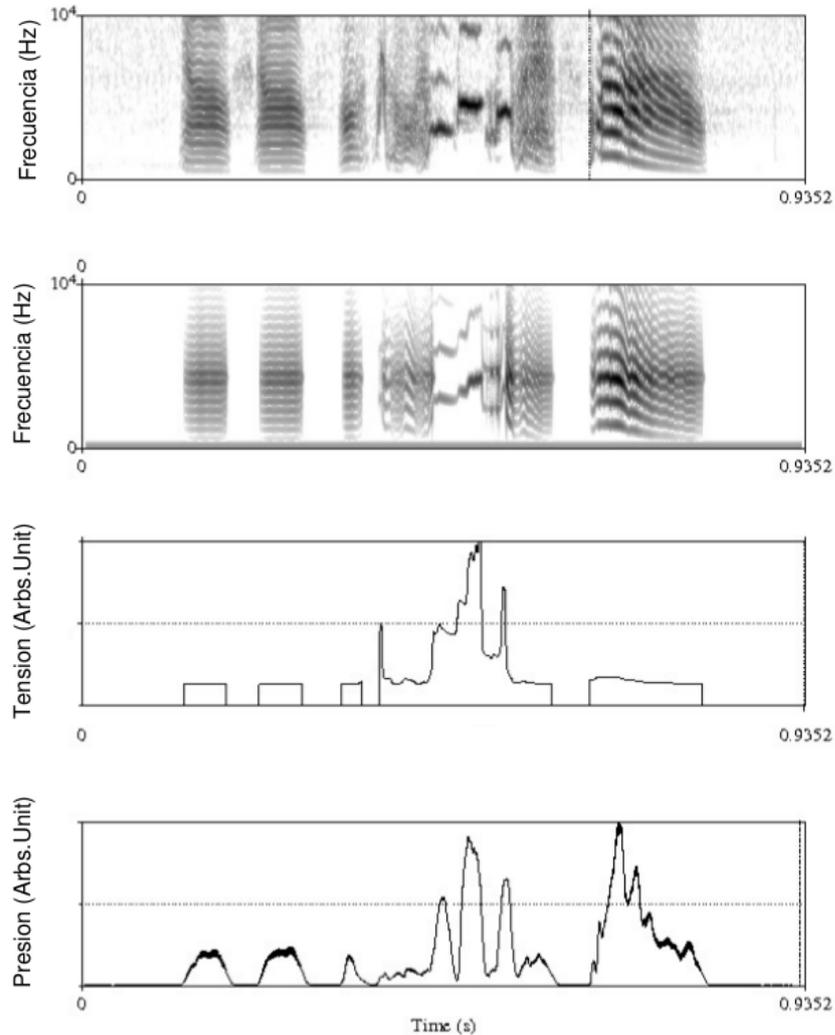


Figura 5.2: En el panel superior se muestra el sonograma de un canto de Diamante mandarín grabado a 44,1 *KHz*; el segundo panel muestra el sonograma del canto sintetizado automáticamente con igual muestreo que el canto original; el tercer y cuarto panel son la señales de tensión y presión respectivamente reconstruidas mediante la automatización del algoritmo.

jación dentro de una caja acústica permitiendo controlar la ubicación del ave en el plano horizontal. Un posicionador micrométrico vertical controla la altura de un electrodo el cual se inserta dentro del HVC y mide actividad eléctrica neuronal. Finalmente con el ave anestesiada y montada en el dispositivo se le presentan estímulos y se mide la respuesta correspondiente. Para identificar neuronas se-

lectivas en HVc se presentan 3 estímulos, el canto propio (BOS), el canto de un conoespecífico (CON) y su canto en reversa (REV). De esta forma se consideran neuronas selectivas al canto propio a aquellas que presentan actividad al estímulo BOS y no a los estímulos REV y CON. Asimismo para evaluar la pertinencia de la síntesis automática a aquellas neuronas que presentaban selectividad se le presenta como estímulo adicional el canto sintético (SYN). Para la obtención de medidas neuronales fueron presentados los 4 estímulos de forma aleatoria al ave anestesiada con 3 motivos por presentación una cantidad de 25 presentaciones por estímulo mientras simultáneamente se registró actividad neuronal. Con las mediciones luego se estableció un criterio para seleccionar la actividad de neuronas individuales e identificar los instantes donde dispara. La figura 5.3 muestra la respuesta de una neurona en HVc de un ave ante los 4 estímulos, en cada panel se muestra la señal temporal del estímulo sonoro, el conteo de disparos representado por histogramas con una ventana temporal de 20 *ms* y la distribución de disparos a lo largo del tiempo para cada presentación.

Se puede notar como la respuesta al canto conoespecífico y reverso es significativamente menor a la respuesta al canto propio. Asimismo ésta presenta una estructura conservada con un pico de actividad en un instante específico e inhibición posterior en las 3 repeticiones del motivo. En la presentación del sintético automático se puede observar la misma estructura en los mismos instantes pero con menor cantidad de cuentas en los máximos de actividad. Si bien la respuesta no es la misma se observa una respuesta positiva al sintético y significativamente distinta a la nula respuesta del conoespecífico y el reverso. Por lo que este experimento muestra resultados promisorios respecto de la pertinencia de la síntesis automática en términos de respuesta neuronal. Actualmente se están continuando estos experimentos aumentando la cantidad de aves para realizar estadística sobre sus resultados y poder así mostrar de forma robusta la bondad de la síntesis automática.

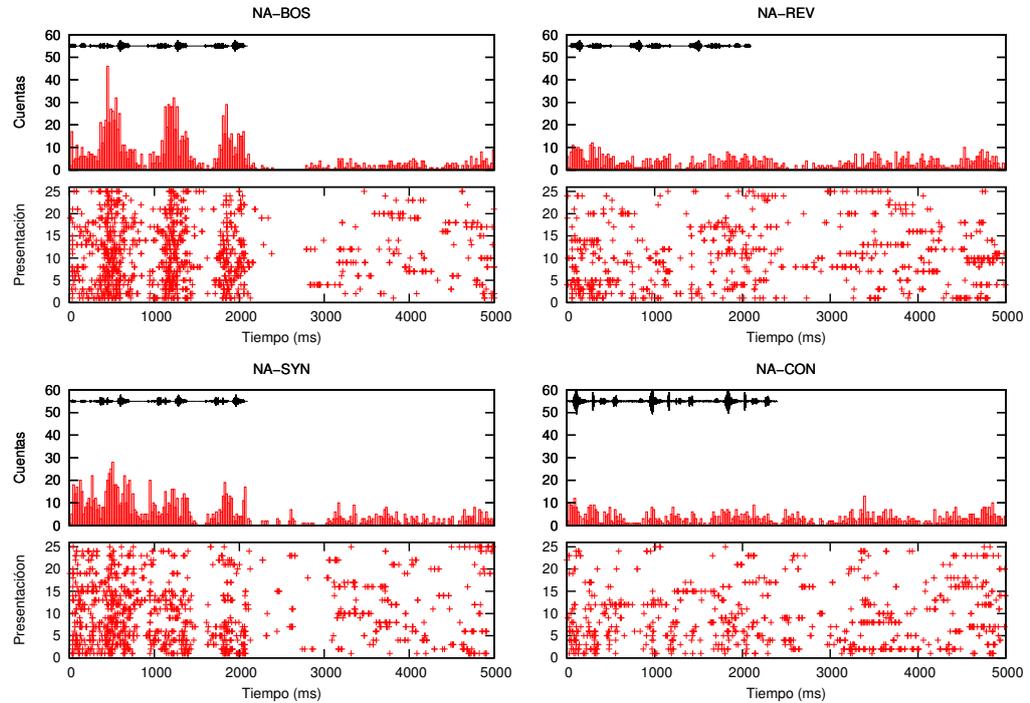


Figura 5.3: Se muestran 4 paneles con la respuesta neuronal de una neurona de HVC a los estímulos BOS, SYN, REV y CON. En cada panel asimismo se observa en la parte superior la señal temporal de las presentaciones sonoras, los histogramas con el conteo de actividad neuronal y debajo los disparos representados por cruces en su respectiva presentación.

5.4. Empleo de la reconstrucción y síntesis automatizada para estudiar la diferencia del contexto social de emisión del canto

Una de las áreas de estudio en la comunidad de canto de aves es aquella que se pregunta si el contexto social de emisión del canto genera variaciones significativas en alguna representación del mismo: acústica, temporal, neuronal, etc. Estudios previos realizados principalmente por el equipo de trabajo de A.Doupe y M.Brainard, han demostrado que existen ciertas diferencias de tiempos silábicos e intersilábicos según el ave ejecuta su canto para cotejo (directamente hacia la hembra, *canto dirigido*) o a modo de práctica (*canto no dirigido*) [Hessler and Doupe, 1999, Brainard and Doupe, 2002]. Así como también en determinados

5.4 Empleo de la reconstrucción y síntesis automatizada para estudiar la diferencia del contexto social de emisión del canto

núcleos relevantes para el aprendizaje (LMAN) se observó actividad diferente según el contexto social de emisión [Hessler and Doupe, 1999]. Por lo tanto, nos propusimos abordar dicha pregunta trasladando la representación a coordenadas motoras. Para ello resulta relevante contar con la herramienta de reconstrucción y síntesis automática de canto que permita procesar grandes cantidades de datos (cantos) para poder estudiar diferencias en la representación motora ante el contexto de emisión. Si bien es un trabajo, en términos de la pregunta científica en proceso, me parece importante contarle como un ejemplo de empleo de la herramienta desarrollada.

A partir de las grabaciones de muchos cantos de un ave realizando canto dirigido (21 cantos) y realizando no dirigido (18 cantos) reconstruimos mediante el código automático los gestos motores. La figura 5.4 muestra la reconstrucción de los gestos de presión y tensión para los cantos dirigidos y no dirigidos en varias repeticiones superpuestos normalizando el tiempo para poder comparar las distintas ejecuciones del motivo.

La conjetura basada en la literatura consistía en que los gestos motores del canto no dirigido debían ser más variables y menos precisos que en el caso de canto dirigido al tratarse en un caso de instancias de práctica y otro de cotejo. Sin embargo como muestra la figura 5.4, no se observan diferencias significativas en los gestos motores en este caso entre ambos contextos de emisión. Es decir, ambos muestran variabilidad de los gestos similares en las sucesivas repeticiones del motivo. Por lo que en esta primera instancia no observamos nada concreto en término de coordenadas motoras, sin embargo consideramos que la herramienta de la síntesis automática y reconstrucción de gestos resultó fundamental para realizar este análisis con esta cantidad de cantos.

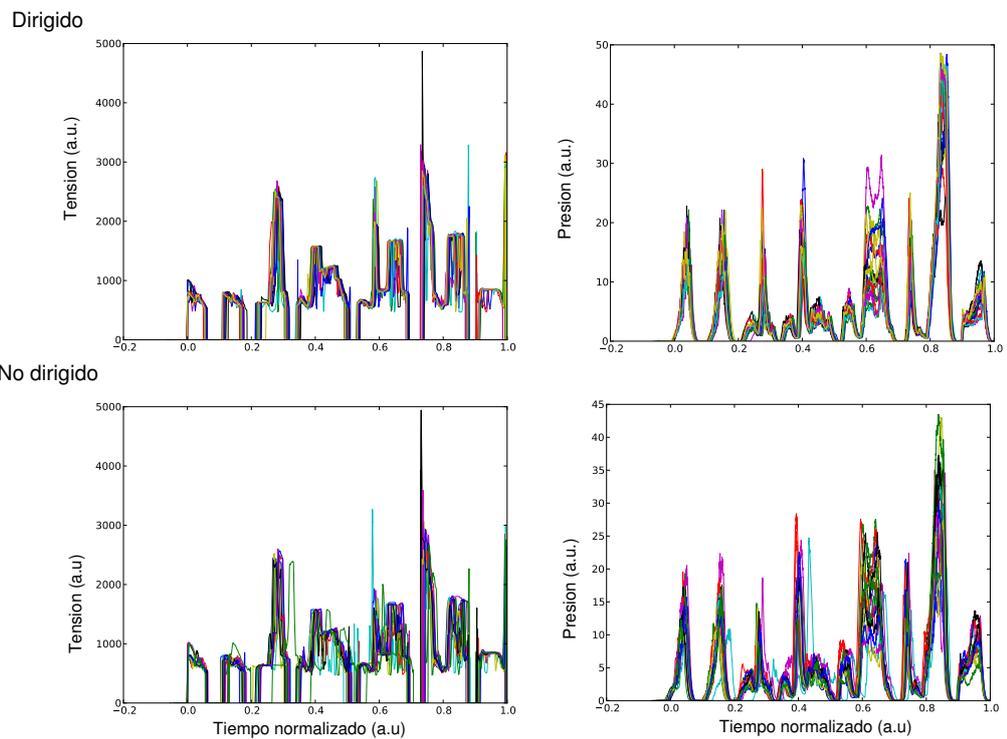


Figura 5.4: Se muestran los gestos reconstituidos automáticamente para un ave en 21 cantos dirigidos y 18 no dirigidos. En la columna de la izquierda se muestra la tensión y en la derecha la presión (arriba dirigido y abajo no dirigido). El tiempo de cada repetición se normalizó para poder comparar todas las repeticiones juntas.

Capítulo 6

Conclusiones

A lo largo de los años de doctorado he tenido la suerte de participar en la parte central de un debate científico donde se revisitaron preguntas que se consideraban respondidas, donde surgieron nuevas preguntas y herramientas de análisis y donde se puso en duda incluso qué es un modelo en el canto de aves. Desde este debate interactuamos con muchos investigadores del área y formamos en el Laboratorio de Sistemas Dinámicos un equipo de trabajo con gente realizando tanto experimentos como modelos capaz de dar el debate a la altura de las circunstancias con convicción, ahínco y honestidad.

Hasta hace unos años atrás la comunidad del canto de ave abocada al estudio de los mecanismos neuronales involucrados en la generación y aprendizaje del canto, consideraba gran parte del sistema del canto resuelto siendo las bases del mismo los trabajos de M. Fee [[Hahnloser et al., 2002](#), [Fee, 2002](#), [Fee et al., 2004](#)]. En estos trabajos M.Fee y co-autores plantearon dos postulados básicos respecto al funcionamiento de sistema motor. El primero plantea un funcionamiento “río abajo” donde las instrucciones motoras del canto se generan inicialmente en el HVc, primer núcleo de la Vía Motora, y luego se comunican “río abajo” en la Vía hasta llegar a las moto-neuronas encargadas del movimiento de los músculos. Asimismo, el HVc conectado directamente con el RA genera actividad neuronal rala pero completa que luego el segundo replica ”uno a uno“ sin generar ningún tipo de actividad propia de la interacción de sus neuronas y así hacia el resto de la vía [[Long et al., 2010](#)]. Esto implica que las instrucciones neuronales que

controlan el canto, según estas conjeturas, suceden en una sola escala temporal que es la que controla HVc y se transmite "río abajo". Mediante esta conjetura M.Fee y co-autores intentaron explicar experimentos donde al enfriar el HVc observaron mayor duración en los motivos del canto [Long and Fee, 2008]. El segundo postulado básico de los trabajos de M. Fee es lo que se dio a conocer como el "modelo del reloj", que proclama que la actividad del HVc es rala y en pulsos equiespaciados de alrededor de 10 *ms* como los "tics" de un reloj sin ninguna correlación con el canto. Por lo tanto las mediciones obtenidas de la actividad de HVc son simplemente algunos "tics" de aquel continuo de disparos que en realidad está ejecutando el HVc durante la generación del canto. Sobre ambos postulados básicos nuestro trabajo reabrió interrogantes y sumado a otros trabajos de distintos investigadores forzó la necesidad de visitar el funcionamiento del sistema motor.

La base de nuestro trabajo consiste plantear la representación del canto en coordenadas motoras como la representación adecuada para comprender cómo codifica el sistema motor. En ese sentido nos basamos en los trabajos de modelado del órgano fonador que durante más de una década se desarrollaron en el Laboratorio de Sistemas Dinámicos. Los trabajos seminales en esa dirección fueron [Gardner et al., 2001, Laje et al., 2002] donde plantearon la posibilidad de realizar modelos de baja dimensión de carácter predictivo capaces de describir la dinámica altamente no lineal que presenta el órgano fonador aviar (siringe). Luego distintos trabajos fueron comprobando la pertinencia de este tipo de modelado en términos de distintas características. Por un lado, en [Sitt et al., 2008] se comprobó que el sonido obtenido sintéticamente mediante el modelo recupera las características acústicas típicas de los sonidos del Diamante mandarín en términos de relación entre frecuencia fundamental y contenido espectral. Luego en [Perl et al., 2011] comprobamos que el modelado era pertinente en término de la reconstrucción de los gestos motores, ya que al sintetizar cantos completos obtuvimos los gestos motores con los que se controla el modelo con una alta correlación con las señales fisiológicas que emplea el ave para controlar la siringe. Finalmente en [Amador et al., 2013] comprobamos la pertinencia en términos de actividad neuronal, ya que observamos que neuronas fuertemente selectivas en HVc presentan respuesta similar al canto propio y al canto sintético al ser presentados como

estímulos auditivos. De esta forma construimos confianza sobre el modelado como herramienta y empleamos los gestos motores para generar conocimiento sobre cómo codifica en términos de dicha representación el sistema del canto. Aplicando este análisis a una población de 44 neuronas del HVC observamos que los disparos de las mismas no suceden en cualquier instante del canto ni de forma descorrelacionada con éste, sino que suceden en instancias precisas de los gestos motores que llamamos GTE (extremos de los gestos motores) [Amador et al., 2013]. Esta observación es incompatible con los dos postulados que dominaban el área. Por un lado si el HVC codifica instancias motoras entonces éste no resulta un reloj que emite un continuo de disparos que suceden durante todo el canto. Por el otro si los disparos en HVC y los extremos en los gestos motores suceden prácticamente en simultáneo, entonces ese disparo en HVC no es el origen de ese gesto motor, es decir es incompatible con un rol de pre-motor para HVC.

Asimismo trabajos de diversos investigadores fueron dando otros indicios sobre la existencia de mecanismos alternativos a los establecidos en la comunidad. Por un lado en [Ashmore, 2008] los autores demostraron la existencia de conectividad desde los núcleos de la parte inferior del sistema (RAm, PAm y DM) hacia los primeros núcleos (HVC y RA) de la supuesta cadena, abriendo la posibilidad a otro tipo de circuito que no sea "río abajo". Por otro, en [Liu et al., 2013] los autores encuentran que estos mismos núcleos de la parte inferior se encuentran conservados evolutivamente en las aves desde las que no aprenden hasta las que si lo hacen. Asimismo experimentos sobre los núcleos inferiores muestran que estimulando DM en aves que aprenden es posible elicitar sonido sin necesidad de la participación en el circuito de HVC [Vicario and Simpson, 1995, Ashmore, 2008]. Estos resultados son indicios de que existe la posibilidad de otro tipo de circuito en el sistema motor donde el HVC no sea el inicio de la señal pre-motora. En [Goldin et al., 2013] experimentos de enfriamiento de HVC y medición de presión simultánea en Diamantes mandarines demostraron presencia de dinámicas propias de la interacción de más escalas de una escala temporal, por lo que la conjetura de que HVC controla la única escala temporal del sistema motor quedaría en duda también.

El debate por lo tanto se instaló en términos de cuál modelo era adecuado para describir el sistema del canto, si la hipótesis del HVC como un reloj o si

la hipótesis de que el HVc responde a los gestos motores como queda planteado en [Troyer, 2013]. Sin embargo, cabe preguntarse si se trata de modelos realmente capaces de describir el sistema del canto. Es decir, el modelo del reloj plantea efectivamente un modelo de funcionamiento que sugiere que un núcleo controla todo en relación uno a uno con el resto de la cadena omitiendo gran cantidad de mediciones no compatibles con ello. Asimismo dicho núcleo dispara constantemente dando cuenta de que cualquier medición es simplemente alguno de todos los continuos disparos que emite el HVc lo que vuelve como modelo irrefutable. Por otro lado el modelo de los gestos motores es una observación de simultaneidad entre eventos lejanos de la cadena (disparos en HVc y extremos en gestos motores) no representa un modelo integral de funcionamiento. Desde esa perspectiva resulta relevante marcar que el logro de nuestro trabajo radica en realizar una observación cuyo resultado es abrir más preguntas que presentar un modelo de funcionamiento de sistema motor. Hoy en día la comunidad de canto de aves está estudiando modelos de sistema motor donde el HVc cumple un rol más predictivo y modulador de las señales motoras que generador de las mismas que resulta más compatibles con las nuevas observaciones.

En este trabajo he estudiado en detalle el modelo de órgano fonador aviar y desarrollado las mejoras necesarias para obtener síntesis de canto pertinente en términos de diferentes espacios métricos. En particular, he estudiado de forma detallada cuál es el comportamiento de las soluciones del sistema de ecuaciones que describen la dinámica de la siringe en su espacio de parámetros. Con ese conocimiento desarrollé un protocolo en el cual a partir de un canto de Diamante mandarín grabado se reconstruyen, vía el modelo, los gestos motores necesarios para generar sonidos con las características acústicas adecuadas. Esta reconstrucción resulta altamente correlacionada con los gestos motores que emplea el ave para generar el canto y a su vez controla el modelo dando por resultado un canto sintético realista [Perl et al., 2011, Perl et al., 2012]. También en búsqueda de mejoras tímbricas en la síntesis de canto realista desarrollé durante esta tesis modificaciones en el modelado. La dinámica del órgano fonador con un filtro estático modelado como tubos ya se encontraba desarrollada y testeada dando resultados muy positivos en términos de parámetros acústicos como ser frecuen-

cia fundamental y contenido espectral [Amador et al., 2008, Sitt et al., 2008], sin embargo ciertas características tímbricas específicas del canto no se encontraban contemplada en dicho modelo. Para ello implementé un modelo de filtro dinámico basado en la fisiología de tracto vocal superior donde se lo modeló como un tubo (la tráquea), una constricción (la glotis), una cavidad resonante (la OEC) y una abertura (el pico) donde irradia el sonido al exterior. De esta forma el filtro es capaz de variar en el tiempo acompañando las necesidades acústicas de cada sílaba y asimismo contempla las tres resonancias más relevantes observadas en los sonogramas del canto, dos del tubo y una de la cavidad. Por otro lado obtuve una mejora tímbrica en la síntesis del canto gracias a la inclusión de ruido en el parámetro que da cuenta de la actividad muscular del vS que le aportó cierta rugosidad al sonido [Perl et al., 2011]. Dichas mejoras tímbricas fueron testeadas en términos de respuesta en neuronas altamente selectivas de HVc. Así estudiamos la existencia de un mecanismo de sintonización de la respuesta de dichas neuronas ante modificaciones de los parámetros estáticos que incluyen mejoras tímbricas (la presencia de ruido en el parámetro y la presencia de la tercer resonancia del resonador). Encontramos que para los valores donde se anula la presencia de las mejoras tímbricas, la respuesta resulta nula y asimismo hallamos valores óptimos donde la respuesta es máxima [Amador et al., 2013].

La similar respuesta de las neuronas selectivas al canto sintético y al canto propio resultó un gran logro para el desarrollo de los modelos desde el punto de vista operacional y conceptual. Por un lado, abre las puertas a la realización de diversos experimentos y preguntas basándose en un canto sintético controlable totalmente por el experimentador a partir de modificaciones en los gestos motores reconstruidos del ave. En esa dirección realizamos trabajos de bio-protética donde el ave enmudecida controla con sus señales fisiológicas a un dispositivo electrónico capaz de integrar el modelo y da por resultado un canto sintético realista controlado en tiempo real por el ave muda [Arneodo et al., 2012]. De esta forma incluyendo modificaciones en los gestos motores que induzcan modificaciones en el canto sintético se pueden realizar estudios de realimentación auditiva alterada. Por otro lado resulta un logro conceptual ya que la reducción de dimensionalidad que propone la realización de modelos sencillos que describen sistemas complejos no resulta obvia en todos los casos. En nuestro trabajo hemos encontrado

una forma de "preguntarle" al sistema que estudiamos cuándo la reducción del sistema resulta adecuada y cuándo exagerada. Es decir, la respuesta neuronal de HVc nos mostró que la reducción del sistema a un oscilador (la siringe)+ un tubo (la tráquea) era insuficiente para elicitación de respuesta, pero que la reducción planteada como un oscilador (siringe) + un tubo (tráquea) + una constricción (glotis) + un resonador (OEC) + una abertura (pico) entonces si era suficiente. Lo que resulta conceptualmente rico tratándose de un sistema biológico donde la dimensionalidad real del sistema es incontable.

Finalmente al tratarse de un procedimiento algorítmico que conduce desde la grabación del canto hasta la obtención de los gestos motores reconstruidos y la síntesis de canto realista abordamos la tarea de automatizar el algoritmo. Por un lado para acelerar los procesos de elaboración de cantos sintéticos y reconstrucción de gestos para abordar mayor cantidad de cantos por proyecto permitiendo realizar preguntas cuya respuesta sea estadística. Por el otro para establecer como una buena representación para el análisis del sistema del canto las coordenadas motoras acercándole a la comunidad de canto de ave una herramienta que permita obtener los gestos motores sin requerir de una capacitación específica. En esa dirección se obtuvo un conjunto de códigos en C y Python capaces de realizar la tarea de forma automática. Actualmente se está analizando en términos de respuesta neuronal la pertinencia del canto sintético obtenido automáticamente con resultados promisorios en las primeras aves experimentadas. Asimismo surgen varios proyectos de aplicación de la automatización donde el procesamiento de grandes datos abre la posibilidad de abordar ciertas preguntas. Por un lado estudiar a partir de los gestos motores si existe alguna diferencia significativa entre el canto dirigido y el no dirigido, estudio que se encuentra en proceso y que los primeros resultados no arrojaron nada clarificador. Por el otro, analizar sistemáticamente los gestos motores de grandes cantidades de aves de diferentes colonias estudiando si existe alguna impronta de los dialectos en los gestos motores.

Appendices

Apéndice A

Figuras Suplementarias al Capítulo 4

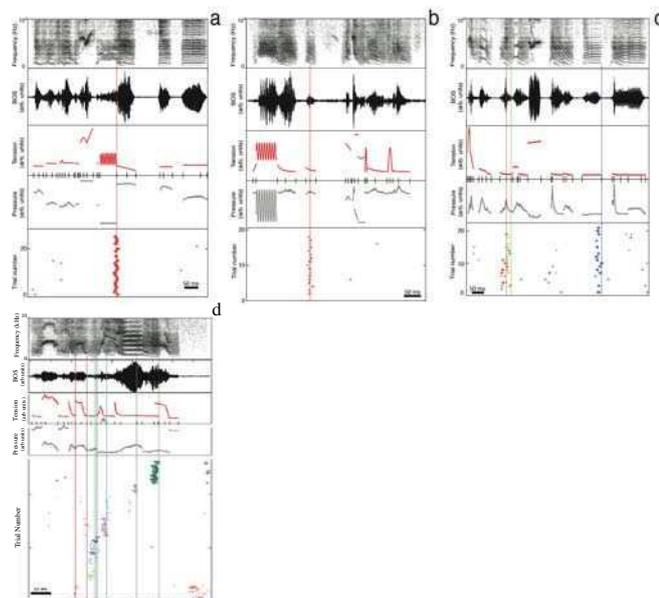


Figura A.1: Se presentan los cantos de las 4 aves restantes para el análisis de neuronas proyectoras en aves dormidas. Cada figura de a)-d) presenta en el panel superior el sonograma del canto, debajo la señal temporal de la presión sonora, luego la tensión y presión reconstruidos y finalmente los disparos neuronales acumulados en los distintos ensayos del estímulo. Las barras negras verticales debajo del panel de tensión muestra la ubicación de los GTE en cada caso.

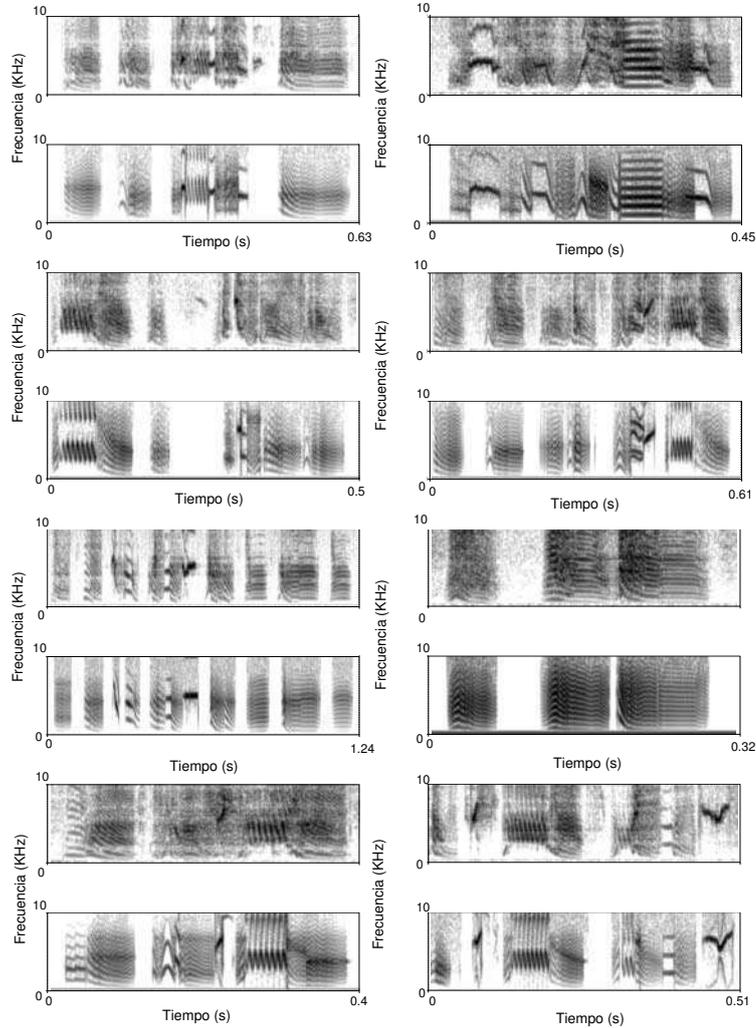


Figura A.2: Se muestran 8 sonogramas de cantos grabados de Diamante mandarín, en la parte superior de cada panel con el correspondiente sonograma del canto sintético obtenido con el modelo en la parte inferior de cada panel. Las 8 grabaciones corresponden a cantos empleados en los experimentos tanto en aves dormidas como en aves despiertas. Notar la variedad acústica de las distintas sílabas y como son recuperadas por el canto modelado.

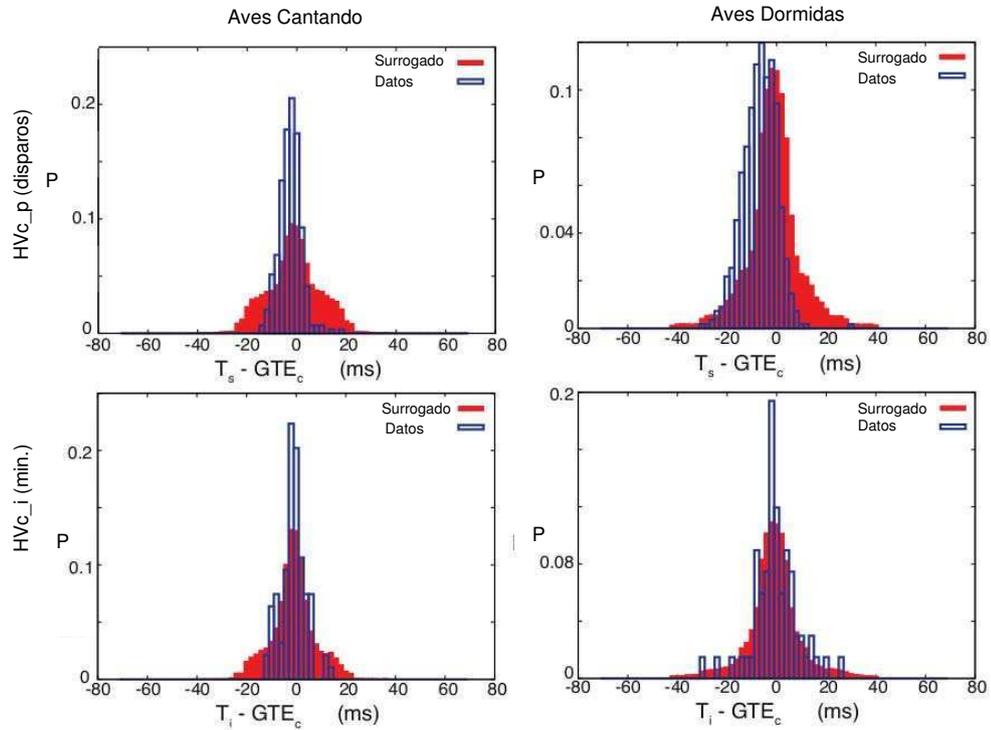


Figura A.3: La distribución de diferencias de tiempos entre los disparos de las neuronas fásicas (T_s) y el GTE más cercano (GTE_c) graficado de manera superpuesta con la distribución de diferencia de tiempos entre poblaciones surrogadas de disparos generados aleatoriamente (respetando la estructura de disparo de los datos) y el GTE_c . En el panel de arriba a la izquierda se muestran los gráficos para las mediciones en aves despiertas y arriba a la derecha para las mediciones en aves dormidas. En los paneles inferiores se muestran las distribuciones de diferencias de tiempo entre los instantes de los mínimos de la actividad reconstruida de las interneuronas (T_i) y el GTE_c . Superpuestas asimismo con la distribución de la diferencia de tiempos entre los mínimos generados aleatoriamente conservando la estructura de los datos originales. En el panel de abajo a la derecha se muestran los gráficos para aves dormidas y en el de la izquierda para aves despiertas. Para la realización de las poblaciones surrogadas generadas se tomaron 1000 repeticiones aleatorias de cada población de igual medida que los datos, se sumaron y luego se normalizaron.

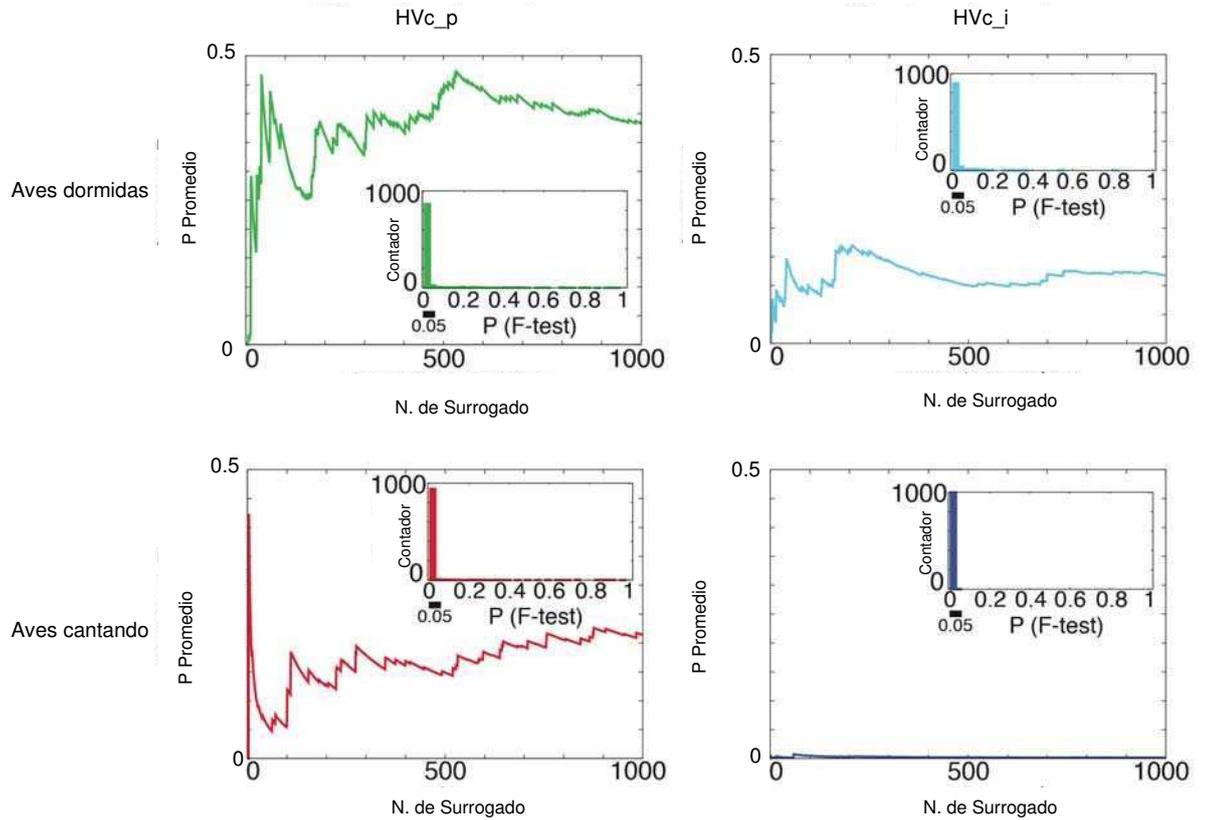


Figura A.4: Los paneles muestran para las poblaciones de mínimos de interneuronas (HVc_i) y para los disparos de las neuronas proyectoras (HVc_p) el análisis estadístico realizado para evaluar la significancia de los resultados. Sobre cada población surrogada (1000), en cada caso, se computó un F-test entre la distribución de tiempos obtenida con los datos y la distribución obtenida con las poblaciones aleatorias. La probabilidad promedio $p_{promedio}$ de que ambas poblaciones provengan de una misma distribución se grafica en cada panel en función del número de población surrogada. Se observa asimismo un histograma de cómo se distribuyen las probabilidades p en cada caso para las 1000 poblaciones surrogadas. Se observa como en los 4 casos la probabilidad p en menor que 0.05 comprobando la significancia de los datos respecto de poblaciones aleatorias.

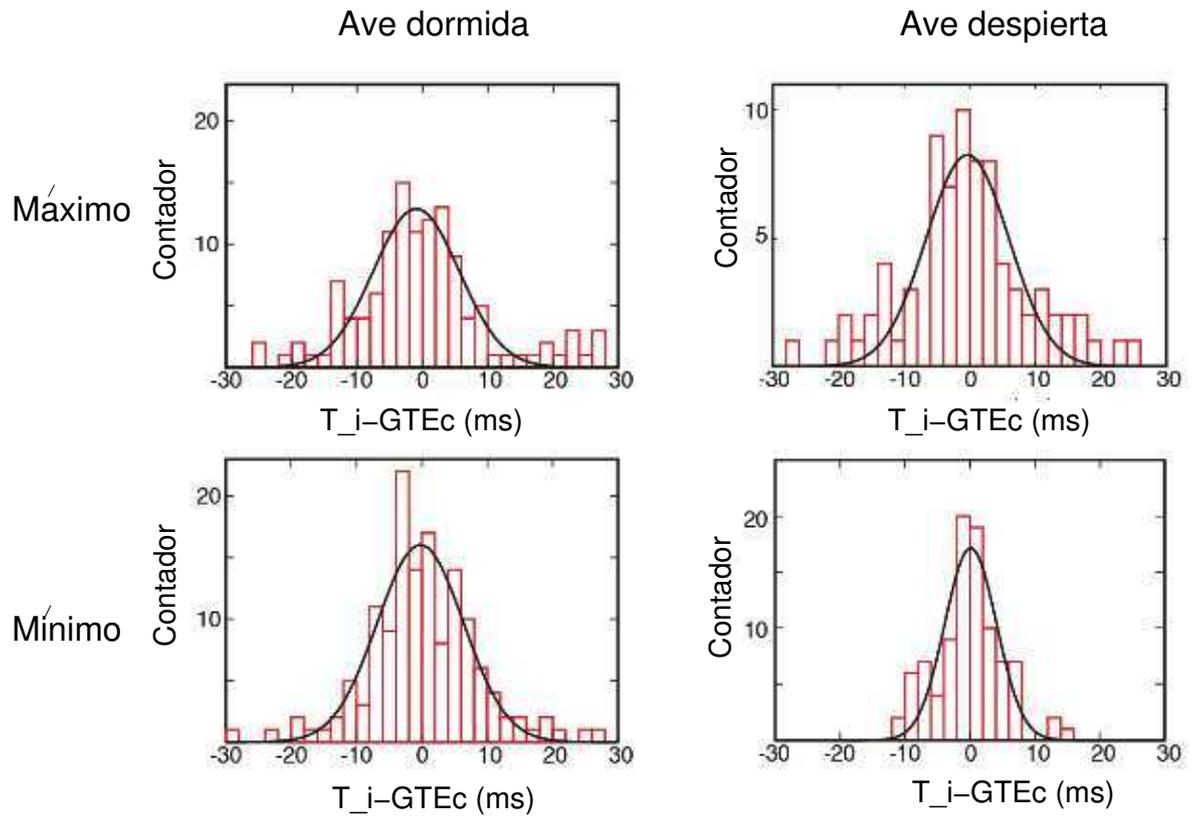


Figura A.5: Los paneles superiores muestran las distribuciones de las diferencias de tiempos entre los GTE y los máximos de actividad de las interneuronas para las aves despiertas y dormidas. En los paneles inferiores se muestra lo mismo pero midiendo la distancia entre los GTE y los mínimos de actividad. Se observa superpuesto sobre cada distribución la campana Gaussiana que mejor la ajusta. Notar cómo las distribuciones de los máximos presentan colas más extendidas que las distribuciones de los mínimos.

Apéndice B

Código en C

Programa escrito en C que computa la SFTF y aplica filtrados para todo un canto completo. La salida del mismo es una serie temporal de frecuencias fundamentales con el mismo muestreo que el canto grabado.

```
//para compilar necesita realft.c four1.c nrutil.c y nrutil.h  
integrando_void.c  
// compilar gcc computa_fundamental_I.c -lm -o g realft.c four1.c  
nrutil.c  
#include <stdio.h>  
#include <stdlib.h>  
#include <math.h>  
#include <string.h>  
#define N_NAME1AUX 50  
  
//-----  
  
int main(int argc, char *argv []) {  
  
//algunas cosas previas al programa.  
    int size, filtro, pow;  
    int largo;  
    int largosobre2;  
    char *entrada;  
  
    if(argc < 3){  
        printf("Debe ingresar tamaño del segmento y el nombre del
```

```
        segmento_"\n");
    return 1;
}

size = atoi(argv[1]);
entrada = argv[2];

largosobre2 = 1024;
largo = 2 * largosobre2;

    int m,h,i,j,k,l,ikp, pasos, sample_max;
    double sampling = 44100.0;
    float data[largo], datareal[largosobre2];
    float dataux[largo];
    float power[largo];
    float frecuencia[largosobre2];
    float serie_fft[largosobre2];
    float serie_temp[size];
    float envuelve[size];
    float sigma=5./1000.*sampling, gaussiana[largosobre2];
    float datareal_filtrada[largosobre2], datareal_filtrada_ant[
        largosobre2];
    float nada;

    int contadormax[size];
    float maximo[largosobre2], freq_max[largosobre2],max, freq;

FILE *ptr;

ptr=fopen(entrada,"r");

    for(m=0;m<size;m++){
        fscanf(ptr,"%f",&serie_temp[m]);
    }

fclose(ptr);

//-----arranca a calcular la SFTF para cada
    tramo de tamaño largosobre2-----//
```

```

for (l=0;l<largosobre2*0.5;l++)    printf(" %d\t %g\t %f\n" ,l ,l
    /44100.,0.);

for (l=largosobre2*0.5;l<size-largosobre2*0.5;l++){

//-----prepara los datos para hacer FTT, le agrego la gaussian
    window con algun sigma-----//
    for (m=-largosobre2*0.5;m<largosobre2*0.5;m++) {
        datareal[m+1] = serie_temp[m+1*1];
        gaussiana[m+1] = exp(-(m+1)*(m+1)/(sigma*sigma));
        datareal_filtrada_ant[m+1] = datareal[m+1]*gaussiana[m+1];
    }

for (m=0;m<largosobre2;m++) {datareal_filtrada [m+1] =
    datareal_filtrada_ant[m+1-512];}

    int isign=1;

    realft(datareal_filtrada , largosobre2 , isign);

j=1;
k=0;
power[1]= datareal_filtrada [1];
power[largosobre2] = datareal_filtrada [2];

for (i=3;i<largosobre2;i+=2){
power[i]=datareal_filtrada [2*i]*datareal_filtrada [2*i]+
    datareal_filtrada [2*i+1]*datareal_filtrada [2*i+1];}

for (i=1;i<largosobre2;i+=2) {
//gaussiana en torno a la fundamental que toca

    serie_fft[k] = sqrt(power[i]);

```

Cap.B Código en C

```
    frecuencia[k] = (i/(1.*largosobre2))*sampling;

    k++;
}

//----- filtros sobre ls STFT-----//
contadormax[1]=0;

for(i=1;i<largosobre2;i++){
    // serie_fft > 12000 al final, es el criterio de superar un umbral
    // . Parametro importante.
    if((serie_fft[i]>serie_fft[i+1])&&(serie_fft[i]>serie_fft[i-1])&&(
        frecuencia[i]>450.)&&(frecuencia[i]<7000.)&&(serie_fft[i
        ]>12000.)){

        frec_max[contadormax[1]] = frecuencia[i];
        contadormax[1]++;
    }
    else{frec_max[contadormax[1]]= 0.;}
}
printf(" %d\t %g\t %f\n",1,(1)/44100.,frec_max[0]);

}
return 0;
}
```

Programa en C que codifica la reconstrucción automática de los gestos dada una señal temporal de frecuencias fundamentales.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <string.h>
#include "finch_void.c"

// gcc sintetiza_III.c rk4.c -lm -o sintetiza
// en la carpeta que este rk4 y el OEC.new.dat

void finch(int size, char *salida, char *envolvente);
```

```

int main(int argc, char *argv[]) {
    int size_pitch_list;
    char *entrada; char *salida; char *envolvente;

    if(argc < 3){
        printf("Debe el nombre del archivo donde está el pitch_list y
            despues el nombre del archivo de salida del beta con le gesto
            de beta y el file con la envolvente\n");
        return 1;
    }

    entrada = argv[1];
    salida = argv[2];
    envolvente = argv[3];

    size_pitch_list = filesize(entrada, 3)-1;

    // printf("%d\n", size_pitch_list);

        int i,j,k, size_song;
        double nada;

        double chi_freq[12500], chi_freq_min;

    double datos1[12500];
double datos2[12500];
double datos3[12500];
double datos4[12500];
        double aproximacion1[size_pitch_list], aproximacion_ant;
double aproximacion2[size_pitch_list];
double aproximacion3[size_pitch_list];
double aproximacion4[size_pitch_list];
double pitch_time[size_pitch_list];
double pitch_value[size_pitch_list];
        FILE *ptr1, *ptr2, *ptr3;

ptr1=fopen("OEC.new.dat", "r");

```

```
ptr2 = fopen(entrada, "r");
ptr3 = fopen(salida, "w");

//levanta de la base de corridas del modelo alfa, beta, f0 y sci

for(i=0;i<12500;i++){
fscanf(ptr1, "%g_%g_%g_%g",&datos1[i],&datos2[i],&datos3[i],&
datos4[i]);
}
fclose(ptr1);

//carga los valores de f0 experimentales de un pitch list

for(i=0;i<size_pitch_list;i++){
fscanf(ptr2, "%g\t%g\t%g\n",&nada, &pitch_time[i],&pitch_value[i
]);
}

fclose(ptr2);

//—————busca el beta y el alfa correspondientes
frecuencia fundamental (fijando el alfa en -0.15)

for(i=0;i<size_pitch_list;i++){
chi_freq_min=10000.;
aproximacion1[i]=0; //alfa
aproximacion2[i]=0; //beta
aproximacion3[i]=0; //frecuencia
aproximacion4[i]=0; //sci

if(pitch_value[i]>10){

for(k=0;k<12500;k++){
if(fabs(datos3[k]-pitch_value[i])<chi_freq_min && (datos1[k]
== -0.15)){
chi_freq_min = fabs(datos3[k]-pitch_value[i]);
```

```

        aproximacion1 [i]=datos1 [k]; aproximacion2 [i]=datos2 [k];
            aproximacion4 [i]=datos4 [k]; aproximacion3 [i]=datos3 [k];
        }
    }
}
else{ aproximacion1 [i]=0.15; aproximacion2 [i]=0.15; aproximacion4 [i
]=0.; aproximacion3 [i]=0.;}
}

for(i=0;i<size_pitch_list;i++) fprintf(ptr3, " %d g\t %d g\n", pitch_time [
i], aproximacion2 [i]);
fclose(ptr3);
finch(size_pitch_list, salida, envolvente);
}
//-----

int filesize(char name[], int ncol){

    int i, size = 1;
    char dato[20];
    FILE *datos;

    datos = fopen(name, "r");

    if(datos==NULL) {
        printf("Falta _archivo_ %s ... \n Saliendo _del_ sistema\n\n", name);
        exit(0);
    }

    while(!feof(datos)){
        for(i=1; i<=ncol; i++) fscanf(datos, "%s", &dato);
        size++;
    }

    fclose(datos);
    return(size-1);
}

```

Cap.B Código en C

```
}
```

Bibliografía

- [Amador et al., 2008] Amador, A., Goller, F., and Mindlin, G. B. (2008). Frequency modulation during song in a suboscine does not require vocal muscles. *J. Neurophysiol*, 99:2383–2389. 14, 20, 26, 29, 32, 103
- [Amador et al., 2013] Amador, A., Perl, Y. S., Mindlin, G. B., and Margoliash, D. (2013). Elemental gestures dynamics are encoded by song premotor cortical neurons. *Nature*. 8, 11, 58, 63, 69, 77, 87, 100, 101, 103
- [Arneodo et al., 2012] Arneodo, E., Perl, Y., F, F. G., and GB, G. M. (2012). Prosthetic avian vocal organ controlled by a freely behaving bird based on a low dimensional model of the biomechanical periphery. *PLoS Comput Biol*, 8(6):e1002546. 103
- [Arneodo and Mindlin, 2009] Arneodo, E. M. and Mindlin, G. B. (2009). Source-tract coupling in birdsong production. *Physical Review E*, 79(6):061921. 23, 49
- [Arneodo et al., 2011] Arneodo, E. M., Perl, Y. S., and Mindlin, G. B. (2011). Acoustic signatures of sound source-tract coupling. *Physical Review E*, 83(4):041920. 23
- [Arrowsmith, 1990] Arrowsmith, D. K. (1990). *An introduction to dynamical systems*. Cambridge University Press. 36
- [Ashmore, 2008] Ashmore (2008). Bottom-up activation of the vocal motor fore-brain by the respiratory brainstem. *The Journal of Neuroscience*, 28(10):2613–2623. 8, 9, 101

- [Bottjer et al., 1984] Bottjer, S., Miesner, E. A., and Arnold, P. A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine bird. *Science*, 224:901–903. 5
- [Brainard and Doupe, 2002] Brainard, M. and Doupe, A. (2002). What birdsong teaches us about learning? *Nature*, 417:351–358. 3, 6, 7, 96
- [Brainard and Doupe, 2000] Brainard, M. S. and Doupe, A. J. (2000). Interruption of a basal ganglia–forebrain circuit prevents plasticity of learned vocalizations. *Nature*, 404(6779):762–766. 5
- [Casey and Gaunt, 1985] Casey, R. M. and Gaunt, A. S. (1985). Theoretical models of the avian syrinx. *Journal of theoretical biology*, 116(1):45–64. 17
- [Chi and Margoliash, 2001] Chi, Z. and Margoliash, D. (2001). Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron*, 32:899–910. 4, 7
- [Daley and Goller, 2004] Daley, M. and Goller, F. (2004). Tracheal length changes during zebra finch song and their possible role in upper vocal tract filtering. *Journal of neurobiology*, 59(3):319–330. 23, 53
- [Dave and Margoliash, 2000] Dave, A. S. and Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*, 290(5492):812–816. 57, 85
- [Doupe and Kuhl, 1999] Doupe, A. and Kuhl, O. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22:567–631. 2
- [Elemans et al., 2003] Elemans, C. P., Larsen, O. N., Hoffmann, M. R., and van Leeuwen, J. L. (2003). Quantitative modelling of the biomechanics of the avian syrinx. *ANIMAL BIOLOGY-LEIDEN-*, 53:183–193. 16
- [Fee, 2002] Fee, M. S. (2002). Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implication for function. *J. Comp. Physiol. A*, 188:829–839. 13, 19, 99

- [Fee et al., 2004] Fee, M. S., Kozhevnikov, A. A., and Hahnloser, R. H. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N.Y. Acad. Sci.*, 1016:153–170. 7, 84, 99
- [Fee et al., 1998] Fee, M. S., Shraiman, B., Pesaran, B., and Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature*, 395(6697):67–71. 27
- [Fiete et al., 2004] Fiete, I. R., Hahnloser, R. H., Fee, M. S., and Seung, H. S. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *Journal of neurophysiology*, 92(4):2274–2282. 59, 83
- [Fletcher, 1992] Fletcher, N. H. (1992). *Acoustic systems in biology*. Oxford University Press Oxford. 52
- [Fletcher et al., 2006a] Fletcher, N. H., Riede, T., and Suthers, R. A. (2006a). Model for vocalization by a bird with distensible vocal cavity and open beak. *The Journal of the Acoustical Society of America*, 119:1005. 22, 23, 50, 52, 53
- [Fletcher et al., 2006b] Fletcher, N. H., Riede, T., and Suthers, R. A. (2006b). Model for vocalization by a bird with distensible vocal cavity and open beak. *The Journal of the Acoustical Society of America*, 119:1005. 50
- [Gardner et al., 2001] Gardner, T., Cecchi, G., Magnasco, M., Laje, R., and Mindlin, G. B. (2001). Simple motor gestures for birdsongs. *Phys. Rev. Lett.*, 87:208101. 26, 29, 33, 48, 49, 100
- [Gardner and Magnasco, 2006] Gardner, T. J. and Magnasco, M. O. (2006). Sparse time-frequency representations. *Proceedings of the National Academy of Sciences*, 103(16):6094–6099. 88
- [Goldin et al., 2013] Goldin, M. A., Alonso, L. M., Alliende, J. A., Goller, F., and Mindlin, G. B. (2013). Temperature induced syllable breaking unveils nonlinearly interacting timescales in birdsong motor pathway. *PLOS ONE*, 8:e67814. 101

- [Goller and Cooper, 2004] Goller, F. and Cooper, B. (2004). Peripheral motor dynamics of song production in zebra finch. *Ann. N.Y. Acad. Sci.*, 59:1016:130. 45
- [Goller and Larsen, 1997] Goller, F. and Larsen, O. N. (1997). A new mechanism of sound generation in songbirds. *Proc. Natl. Acad. Sci.*, 94:14787–14791. 19
- [Goller et al., 2004] Goller, F., Mallinckrodt, M. J., and Torti, S. D. (2004). Beak gape dynamics during song in the zebra finch. *Journal of neurobiology*, 59(3):289–303. 23
- [Goller and Suthers, 1996a] Goller, F. and Suthers, R. (1996a). Role of syringeal muscles in controlling the phonology of bird song. *J Neurophysiol*, 76:287–300. 20, 43
- [Goller and Suthers, 1996b] Goller, F. and Suthers, R. (1996b). Role of syringeal muscles in in gating airflow and sound production in singing brown thrashers. *J Neurophysiol*, 75:867–876. 20, 21
- [Guckenheimer and Holmes, 1997] Guckenheimer, J. and Holmes, P. (1997). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer-Verlag. 36
- [Hahnloser et al., 2002] Hahnloser, R. H. R., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419:65–70. 6, 7, 11, 64, 83, 84, 99
- [Hartley and Suthers, 1989] Hartley, R. S. and Suthers, R. A. (1989). Airflow and pressure during canary song: evidence for mini-breaths. *J. Comp. Physiol.*, 165:15–26. 70
- [Hessler and Doupe, 1999] Hessler, N. A. and Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forabrain. *Nature Neuroscience*, 2(3). 96, 97
- [Jarvis, 2004] Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016(1):749–777. 3

- [Jensen et al., 2007] Jensen, K. K., Cooper, B. G., Larsen, O. N., and Goller, F. (2007). Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proceedings of the Royal Society B: Biological Sciences*, 274(1626):2703–2710. 34
- [Katz and Gurney, 1981] Katz, L. C. and Gurney, M. E. (1981). Auditory responses in the zebra finch’s motor system for song. *Brain research*, 221(1):192–197. 55
- [Kelley and Nottebohm, 1979] Kelley, D. B. and Nottebohm, F. (1979). Projections of a telencephalic auditory nucleus–field l–in the canary. *Journal of Comparative Neurology*, 183(3):455–469. 5
- [Keynes et al., 2001] Keynes, R. D., Aidley, D. J., and Huang, C. L.-H. (2001). *Nerve and muscle*. Cambridge Univ Press. 61
- [Kinsler et al., 1982] Kinsler, L., Frey, A., and Coppens, A. (1982). *Fundamentals of acoustics*. John Wiley & Sons, Inc., New York. 52
- [Konishi, 1965] Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift für Tierpsychologie*, 22:770–783. 2
- [Konishi, 1978] Konishi, M. (1978). Auditory environment and vocal development in birds. In *Perception and experience*, pages 105–118. Springer. 2
- [Kozhevnikov and Fee, 2007] Kozhevnikov, A. A. and Fee, M. S. (2007). Singing-related activity of identified hvc neurons in the zebra finch. *Journal of neurophysiology*, 97(6):4271–4283. 64
- [Laje et al., 2002] Laje, R., Gardner, T. J., and Mindlin, G. B. (2002). Neuromuscular control of vocalizations in birdsong: A model. *Phys. Rev. E*, 65:051921. 26, 29, 33, 50, 100
- [Larsen and Goller, 1999] Larsen, O. N. and Goller, F. (1999). Role of syringeal vibrations in bird vocalizations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1429):1609–1615. 27

- [Larsen and Goller, 2002] Larsen, O. N. and Goller, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *J. Exp. Biol.*, 205. 21
- [Lim et al., 2012] Lim, Y., Shinn-Cunningham, B., and Gardner, T. J. (2012). Sparse contour representations of sound. *Signal Processing Letters, IEEE*, 19(10):684–687. 88, 89
- [Liu et al., 2013] Liu, W.-c., Wada, K., Jarvis, E., and Nottebohm, F. (2013). Rudimentary substrates for vocal learning in a suboscine. *Nature communications*, 4. 10, 101
- [Long and Fee, 2008] Long, M. A. and Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, 456:189–194. 100
- [Long et al., 2010] Long, M. A., Jin, D. Z., and Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature*, 468(7322):394–399. 99
- [Luo et al., 2001] Luo, M., Ding, L., and Perkel, D. J. (2001). An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *The Journal of Neuroscience*, 21(17):6836–6845. 6
- [Margoliash, 1983] Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *The Journal of Neuroscience*, 3(5):1039–1057. 5, 56, 67, 85
- [Margoliash, 1986] Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *The Journal of neuroscience*, 6(6):1643–1661. 56, 67, 85
- [Margoliash and Fortune, 1992] Margoliash, D. and Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch’s hvc. *The Journal of neuroscience*, 12(11):4309–4326. 5, 57, 67, 85
- [Marler, 1967] Marler, P. (1967). Animal communication signals. *Science*. 2

- [Marler, 1970] Marler, P. (1970). A comparative approach to vocal learning: song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology.*, 71:1. 2
- [McCasland and Konishi, 1981] McCasland, J. S. and Konishi, M. (1981). Interaction between auditory and motor activities in an avian song control nucleus. *Proceedings of the National Academy of Sciences*, 78(12):7815–7819. 4, 56
- [Mindlin et al., 2003] Mindlin, G. B., Gardner, T. J., Goller, F., and Suthers, R. (2003). Experimental support for a model of birdsong production. *Phys. Rev. E*, 68:041908. 26, 30, 50, 61
- [Mindlin and Laje, 2005] Mindlin, G. B. and Laje, R. (2005). *The Physics of Birdsong*. Springer, New York. 13, 14, 18, 19
- [Nottebohm et al., 1976] Nottebohm, F., Stokes, T. M., and Leonard, C. M. (1976). Central control of song in the canary, *serinus canarius*. *Journal of Comparative Neurology*, 165(4):457–486. 4, 14
- [Nowicki, 1987] Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature*, 325(6099):53–55. 23
- [Ohms et al., 2010] Ohms, V. R., Snelderwaard, P. C., ten Cate, C., and Beckers, G. J. (2010). Vocal tract articulation in zebra finches. *PloS one*, 5(7):e11923. 50
- [Perl et al., 2011] Perl, Y. S., Arneodo, E. M., Amador, A., Goller, F., and Mindlin, G. B. (2011). Reconstruction of physiological instructions from zebra finch song. *Phys. Rev. E*, 84:051909. 26, 38, 39, 42, 44, 48, 50, 53, 54, 58, 87, 89, 91, 92, 100, 102, 103
- [Perl et al., 2012] Perl, Y. S., Arneodo, E. M., Amador, A., and Mindlin, G. B. (2012). Nonlinear dynamics and the synthesis of zebra finch song. *International Journal of Bifurcation and Chaos*, 22(10). 48, 50, 58, 102

- [Prather et al., 2008] Prather, J., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451:305–310. 8, 57
- [Prather et al., 2009] Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., and Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature neuroscience*, 12(2):221–228. 85
- [Press, 2007] Press, W. H. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press. 78
- [R.A.Suthers et al., 1999] R.A.Suthers, Goller, F., and Pytte, C. (1999). The neuromuscular control of birdsong. *Philos Trans R Soc Lond B Biol Sci.*, 354:927–39. 5
- [Riede et al., 2006] Riede, T., Suthers, R. A., Fletcher, N. H., and Blevins, W. E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl. Acad. Sci.*, 103(14):5543–5548. 13, 23
- [Roberts et al., 2008] Roberts, T. F., Klein, M. E., Kubke, M. F., Wild, J. M., and Mooney, R. (2008). Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *The Journal of Neuroscience*, 28(13):3479–3489. 9
- [Shank and Margoliash, 2008] Shank, S. S. and Margoliash, D. (2008). Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature*, 458(7234):73–77. 57, 58
- [Sitt et al., 2008] Sitt, J. D., Amador, A., Goller, F., , and Mindlin, G. B. (2008). Dynamical origin of spectrally rich vocalizations in birdsong. *Phys. Rev. E*, 78:011905. 14, 15, 17, 20, 26, 33, 34, 35, 100, 103
- [Sitt et al., 2010] Sitt, J. D., Arneodo, E. M., Goller, F., and Mindlin, G. B. (2010). Physiologically driven avian vocal synthesizer. *Phys. Rev. E*, 81:031927. 15, 20, 30, 36, 43

- [Spiro et al., 1999] Spiro, J. E., Dalva, M. B., and Mooney, R. (1999). Long-range inhibition within the zebra finch song nucleus ra can coordinate the firing of multiple projection neurons. *J. Neurophysiol.*, 81:3007–3020. 5
- [Strogatz, 2001] Strogatz, S. (2001). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering (studies in nonlinearity)*. Westview Press, Boulder. 30, 31
- [Suthers, 2001] Suthers, R. A. (2001). Peripheral vocal mechanisms in birds: are songbirds special? *Netherlands J. Zool.*, 51:217–242. 16
- [Suthers and Margoliash, 2002] Suthers, R. A. and Margoliash, D. (2002). Motor control of birdsong. *Curr. Opin. Neurobiol.*, 12:684–690. 43
- [Sutter and Margoliash, 1994] Sutter, M. L. and Margoliash, D. (1994). Global synchronous response to autogenous song in zebra finch hvc. *Journal of neurophysiology*, 72(5):2105–2123. 77
- [Theunissen and Doupe, 1998] Theunissen, F. E. and Doupe, A. J. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus hvc of male zebra finches. *The Journal of neuroscience*, 18(10):3786–3802. 67
- [Thorpe, 1961] Thorpe, W. H. (1961). *Bird-song: the biology of vocal communication and expression in birds*. University Press. 2, 15
- [Titze, 1988] Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.*, 83:1536–1550. 27, 28
- [Titze, 1994] Titze, I. R. (1994). *Principles of voice production*. Prentice Hall Englewood Cliffs. 48
- [Troyer, 2013] Troyer, T. W. (2013). Neuroscience: The units of a song. *Nature*, 495(7439):56–57. 12, 102
- [Vicario and Simpson, 1995] Vicario, D. S. and Simpson, H. B. (1995). Electrical stimulation in forebrain nuclei elicits learned vocal patterns in songbirds. *Journal of neurophysiology*, 73(6):2602–2607. 10, 101

- [Vu et al., 1994] Vu, E. T., Mazurek, M. E., and Kuo, Y.-C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *The Journal of neuroscience*, 14(11):6924–6934. 85
- [Wiggins, 2003] Wiggins, S. (2003). *Introduction to applied nonlinear dynamical systems and chaos*, volume 2. Springer. 36
- [Williams and Vicario, 1993] Williams, H. and Vicario, D. S. (1993). Temporal patterning of song production: participation of nucleus uvaeformis of the thalamus. *Journal of neurobiology*, 24(7):903–912. 85
- [Woolley and Doupe, 2008] Woolley, S. C. and Doupe, A. J. (2008). Social context-induced song variation affects female behavior and gene expression. *PLoS biology*, 6(3):e62. 15
- [Yu and Margoliash, 1996] Yu, A. C. and Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science*, 273:1871–1875. 4