

Tesis Doctoral

Aportes al estudio cuantitativo del lenguaje y la comunicación animal en tres etapas: recepción, elaboración y emisión de un mensaje

Costa, Martín Elías

2014-08-11

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:

Costa, Martín Elías. (2014-08-11). Aportes al estudio cuantitativo del lenguaje y la comunicación animal en tres etapas: recepción, elaboración y emisión de un mensaje. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Costa, Martín Elías. "Aportes al estudio cuantitativo del lenguaje y la comunicación animal en tres etapas: recepción, elaboración y emisión de un mensaje". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2014-08-11.

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 Facultad de Ciencias Exactas y Naturales.

Departamento de Física

**Aportes al estudio cuantitativo del
lenguaje y la comunicación animal en
tres etapas: recepción, elaboración y
emisión de un mensaje.**

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

Martín Elías Costa

Directores de tesis: Dr. Mariano Sigman y Dr. Diego Laplagne

Consejero de estudios: Dr. Ricardo Piegaia

Lugar de trabajo: Laboratorio de Neurociencia Integrativa, Departamento de Física.

Defensa: Buenos Aires, 11 de Agosto de 2014



Abstract

Este trabajo de tesis trata sobre la comunicación oral y está dividida en tres grandes subtemas. El primero de ellos comprende la conversión de ondas de presión en el tímpano (sonido) en conceptos; el segundo aborda la transición mental entre ideas y finalmente el tercero, la emisión de una vocalización. De esta manera cubrimos el recorrido completo que experimenta una señal en un hablante. Cada una de estas tres secciones es un campo en si mismo y lo que se encuentra en este trabajo son pequeños aportes a cada uno de ellos. En el área del proceso de comprensión del habla estudiamos un efecto que se conoce como Efecto de Transformación Verbal (ETV) que se produce frente a la escucha repetida de la misma palabra y mostramos que su dinámica cumple con todas las características de percepción biestable. Esto agrega al ETV a la lista de fenómenos biestables junto con las figuras ambiguas, tonos galopantes, rivalidad binocular, etc. y sugiere modificaciones a los modelos vigentes de reconocimiento de palabras. En el campo de la semántica cuantitativa, construimos un grafo semántico de co-ocurrencias (una medida estadística de la similitud entre palabras a partir de la estructura de sus ocurrencias en texto escrito) y mostramos que el proceso de asociación libre (decir la primera palabra que se nos viene a la cabeza frente a la presentación de una palabra estímulo) puede ser capturado, en esencia, por un proceso difusivo anómalo dentro de dicho grafo. Finalmente, estudiamos la emisión de vocalizaciones, no ya en humanos sino en ratas, analizando la riqueza de su comportamiento vocal con miras a establecerlas como un modelo animal de comunicación vocal. En particular, desarrollamos algoritmos de clasificación automática para sus vocalizaciones ultrasónicas y mostramos que se estructuran de manera probabilística en el tiempo. Con esto se cierra nuestro ciclo de recepción, elaboración y emisión de un mensaje.

Palabras clave

lenguaje - comunicación animal - comprensión de palabra hablada - lingüística cuantitativa - vocalizaciones ultrasónicas

Abstract

The work presented here is a study on oral communication and is subdivided into three big areas. The first one deals with the translation of pressure waves at the eardrums (sound) into concepts in the brain; the second one explores the mental transition between thoughts and finally, the third one investigates vocal output. In this manner, we cover the whole excursion of a signal through a communicating agent. Each one of these stages represents an active and broad field of research. Here, we present our contributions to each one of those areas. In the area of spoken word recognition we study the Verbal Transformation Effect (VTE) which entails a change in the perceived word after repeated presentations. We show that the dynamics of this effect has all the characteristics of bistable perception. This appends the VTE to the list of bistable percepts (together with ambiguous figures, galloping tones, binocular rivalry, etc) and suggests corrections to current models of spoken word recognition. In the field of quantitative semantics, we constructed a semantic graph of word co-occurrence (a statistical measure of similarity between words based on the structure of occurrences in written text) and showed that the process of free association (i.e saying first thing that comes to mind upon the presentation of a stimulus word) can be understood as an anomalous diffusion process in the mentioned graph. Finally, we study the production of vocalizations, not in humans but in the rat. We analyze the richness of their vocal behavior in an attempt to establish them as an animal model for vocal communication. In particular, we developed algorithms for automatic classification of ultrasonic vocalizations and showed that they are probabilistically structured in time. This closes the communication loop: to receive, elaborate and send a message.

Keywords

language - animal communication - spoken word recognition - quantitative
linguistics - ultrasonic vocalizations

A Tata y Ana

Agradecimientos

Nadie es lo que es sin quienes lo rodean; sin la gente a la que quiere. Yo tuve la suerte de estar rodeado de gente genial desde que nací y la astucia para elegir seguir estándolo hoy. A ustedes, que me saben de pocas palabras, quiero decirles simplemente: gracias.

A mis viejos por ser cuna, trampolín y refugio

A mis hermanos porque no soy uno, sino una parte de tres

A la mesa chica a la que quiero infinito, a Flor, Lula, Vico, Paco y ?

A Belu por el estilo

A Kers por la luz

A Emi y Ro por su optimismo sin límites

A Mariano F por los cuentos que vendrán

A los jueves por bancar mi pie de palo

Al Brío por despertarme

Al LNI por las ondas alfa

A los secuaces del LSD: MT, ZK e YS

A los de Roque: Andrew, Rob, Pawel y Yevgeny

A Mariano S por el largo camino

A Diego L por la opera

A Wais por lo feliz

Índice general

Índice de figuras	VII
Índice de cuadros	IX
Abreviaturas	XI
1. Apertura	1
1.1. Recepción, emisión y elaboración	1
1.2. Organización de esta Tesis	3
2. Biestabilidad Lexical	5
2.1. Introducción	5
2.1.1. Sobre el Efecto de Transformación Verbal	8
2.1.2. Sobre la Biestabilidad Perceptual	10
2.2. El experimento de Biestabilidad Verbal	13
2.2.1. Las ventajas del castellano en la construcción de estímulos	13
2.2.2. Paradigma Experimental	16
2.2.3. Resultados: El ETV es un fenómeno biestable	17
2.3. Discusión	18
2.3.1. ¿De dónde vienen las oscilaciones?	20
2.3.2. Control Voluntario	21
2.4. Conclusiones	22
3. Asociación libre de palabras	23
3.1. Introducción	23
3.2. Una métrica para el espacio de palabras	25
3.2.1. Construcción del Corpus	25

ÍNDICE GENERAL

3.2.2. Definición de la métrica entre conceptos	26
3.3. Experimentos de asociación libre	26
3.3.1. Plataforma para experimentación Online	27
3.4. Resultados	28
3.4.1. Fuzzy clustering y proyección de Sammon	28
3.4.2. Los ciclos y el comportamiento difusivo	29
3.5. Modelos difusivos para la asociación libre	32
3.5.1. Modelos de vecindario acotado	33
3.5.2. Modelos libres de escala	35
3.6. Conclusiones	39
4. Vocalizaciones Ultrasónicas en Ratas.	41
4.1. Introducción	42
4.2. Características sonoras	42
4.2.1. Rangos, duraciones y propiedades espectrales	42
4.2.2. Mecanismos de producción	43
4.2.3. Clases de Vocalizaciones	43
4.3. Rol Social	44
4.3.1. Vocalizaciones de 22 KHz	46
4.3.2. Vocalizaciones de 50 KHz	46
4.4. Circuitos Neuronales	47
4.5. Control Voluntario	48
4.6. Respiración	49
4.7. Dispositivo Experimental	49
4.7.1. Hardware	50
4.7.2. Software	50
4.7.3. Protocolo	51
5. Clasificación automática y estructura de vocalizaciones	53
5.1. Estrategia general	53
5.2. Detección y Extracción de propiedades	55
5.2.1. Cálculo del espectro.	56
5.2.2. Segmentación de vocalizaciones	57
5.2.3. Propiedades de la curva	59

ÍNDICE GENERAL

5.2.4. Mapa de Propiedades	61
5.3. Combinación en secuencias	66
5.4. Estructura entre vocalizaciones	73
5.4.1. Composición de categorías en una ráfaga	73
5.4.2. Memoria del sistema	77
5.5. Interacción entre individuos	81
5.6. Conclusiones	82
6. Otros aspectos de la producción vocal en ratas	85
6.1. Introducción	85
6.2. Vocalización y locomoción	85
6.3. Mecanismos de producción	89
6.3.1. Conclusiones	94
6.4. Control Vocal	97
6.4.1. Configuración Experimental	97
6.4.2. Resultados	101
6.4.3. Conclusiones	108
7. Cierre	111
Bibliografía	113

ÍNDICE GENERAL

Índice de figuras

2.1. Ilusión de perspectiva	6
2.2. Cubo de Necker	8
2.3. Chitas ambiguas	10
2.4. Resumen de Biestabilidad Perceptual	14
2.5. Manipulación de los estímulos	15
2.6. Esquema del paradigma experimental	17
2.7. Características biestables en los datos experimentales	19
2.8. Ajuste lineal del par CASA-SACA	20
3.1. Proyección de una trayectoria	30
3.2. Datos Experimentales	31
3.3. Simulaciones para vecindarios acotados	36
3.4. Densidades de probabilidad de los distintos modelos	37
3.5. Simulaciones libres de escala	38
4.1. Espectros de una vocalizaciones ultrasónicas	44
4.2. Categorías de Vocalizaciones	45
4.3. Arena de Grabación	51
4.4. Análisis de los videos	52
5.1. Estrategia general para la categorización de curvas.	54
5.2. Ejemplos de espectrogramas	56
5.3. Ejemplo de extracción de una curva	58
5.4. Histograma de duración de silencios	59
5.5. Distribución espectral de potencia de las curvas	60
5.6. Ejemplo de curva con su regresión	61

ÍNDICE DE FIGURAS

5.7. Mapas de propiedades	63
5.8. Excursiones de vocalizaciones en el espacio de propiedades	64
5.9. Histograma del parámetro B en el ajuste	65
5.10. Histogramas de duración para las nubes F y T	65
5.11. Modelo de Markov	67
5.12. Comparación de secuencias con el modelo	68
5.13. Comparación de clasificaciones	70
5.14. Ejemplo de un FT mal categorizado	71
5.15. Distribución de frecuencias iniciales según secuencias	72
5.16. Tabla de nueva notación	74
5.17. Composición de las ráfagas en función de su largo	75
5.18. Composición de la ráfaga en función de su largo (varias ratas)	76
5.19. Proporción de Categorías en función del orden	78
5.20. Histogramas de Tiempo entre vocalizaciones para dos experimentos	82
6.1. Velocidad y vocalizaciones	86
6.2. Filtros entre Vocalizaciones y velocidad	88
6.3. Mapa de una frecuencia en un instante contra el siguiente	90
6.4. Esquema de un silbato <i>bird-call</i>	90
6.5. Mapa de saltos con rectas dadas por la ec 6.4	92
6.6. Anatomía del tracto respiratorio de la rata	93
6.7. Frecuencia de oscilación en función de la velocidad	95
6.8. Evolución temporal del flujo	96
6.9. Configuración del dispositivo experimental	100
6.10. Entrenamiento de una Rata	102
6.11. Inversión de la regla	103
6.12. Entrenamiento Vocal	105
6.13. Entrenamiento Vocal segunda fase	106
6.14. Histogramas de tiempo de Respuesta según distintas condiciones	107
6.15. Ejemplos de Vocalizaciones	109
6.16. Cambios en los mapas de R6	110

Índice de cuadros

2.1. Lista de sílabas estímulo.	16
2.2. Lista de palabras estímulo.	16
3.1. Lista de palabras semilla en el juego de asociación.	28
5.1. Parámetros de los espectrogramas y umbral para la detección de vocalizaciones en el archivo de audio. Hth1 es el primer umbral de entropía; Hth2 el segundo; WIN es el tamaño de la ventana del espectrograma y STEP el número de paso.	57

ABREVIATURAS

Abreviaturas

ACC Corteza Cingulada Anterior
AIC Criterio de Informacion de Akaike
BIC Criterio de Informacion Bayesiano
CFI Corteza Frontal Inferior

EEG ElectroEncefaloGrafía
ETV Efecto de Transformación Verbal
FM Frecuencia Modulada
GCA Giro Cingulado Anterior
LB Método de Lattice Boltzmann (Cuadrícula de Boltzmann)
LRF Formación Reticular Lateral
LTN Núcleo Tegmental Laterodorsal
NA Núcleo Ambiguo
PAG Sustancia Gris Periacueductal
RFID Identificación de Radiofrecuencia

ABREVIATURAS

1

Apertura

1.1. Recepción, emisión y elaboración

¿Cómo es que un conjunto de sonidos se interpreta como una palabra? ¿Cómo se conecta un significado con otro? ¿Qué habilidades vocales compartimos con otros mamíferos (y cuales no)? Estas tres preguntas se inscriben cada una dentro de una parte del bucle infinito que hace a cualquier sistema recíproco de comunicación. Un agente recibe un estímulo físico, una variación de presión en el aire, una variación de campo electromagnético, una palmada en la espalda. Ese estímulo se decodifica, modifica al agente, que lo elabora, realiza una serie de cálculos y transformaciones y puede o no accionar sobre el ambiente para generar otro estímulo como respuesta a su interlocutor.

Esta tesis es un recorrido por esas tres etapas. Por supuesto que cada una de las etapas, así definidas, es un campo de investigación en si mismo y lo que se cuenta en estas páginas son apenas distintos paseos exploratorios por cada uno de esos campos. Podría decirse que esto se corresponde con una mirada generalista sobre el proceso de comunicación. Este abordaje tiene la desventaja de que ninguno de los temas puede ser analizado a fondo, pero, se anota a favor el brindar una mirada más global e integradora.

La división del proceso en estas tres partes no es novedosa, tiene sus raíces en la interpretación Jamesiana de la acción como acto reflejo.

“I refer to the doctrine of reflex action, especially as extended to the brain. This is, of course, so familiar to you that I hardly need define it. In a general way, all educated people know what reflex action means.

1. APERTURA

It means that the acts we perform are always the result of outward discharges from the nervous centres, and that these outward discharges are themselves the result of impressions from the external world, carried in along one or another of our sensory nerves. Applied at first to only a portion of our acts, this conception has ended by being generalized more and more, so that now most physiologists tell us that every action whatever, even the most deliberately weighed and calculated, does, so far as its organic conditions go, follow the reflex type. There is not one which cannot be remotely, if not immediately, traced to an origin in some incoming impression of sense. There is no impression of sense which, unless inhibited by some other stronger one, does not immediately or remotely express itself in action of some kind. There is no one of those complicated performances in the convolutions of the brain to which our trains of thought correspond, which is not a mere middle term interposed between an incoming sensation that arouses it and an outgoing discharge of some sort, inhibitory if not exciting, to which itself gives rise. The structural unit of the nervous system is in fact a triad, neither of whose elements has any independent existence.”

William James - Reflex Action And Theism 1881

En este esquema, acto reflejo no es un eufemismo; quiere decir literalmente eso. Un reflejo como el de la rodilla, perfectamente entendido en su mecánica. Una neurona sensorial de la rodilla hace sinapsis al mismo tiempo sobre dos neuronas: una motoneurona (que activa el músculo extensor) y una interneurona (que a su vez hace sinapsis inhibitoria sobre una motoneurona que inerva el músculo flexor). El resultado: la pierna se contrae frente a un pequeño golpe. Así, un comportamiento más elaborado deriva de una complejización de la red de interneuronas que se anteponen en el camino del impulso. Un camino esculpido por genes y experiencia.

Esta tesis es, más que nada, una excusa para dar una vuelta en ese bucle. Siempre me llamó un poco la atención como esta idea permea a toda la disciplina neurocientífica y sin embargo raramente se menciona el ciclo completo. Por supuesto, es una omisión por necesidad y no por ignorancia. Si uno se va a poner a estudiar en detalle los mecanismos de la memoria, es razonable que omita hacer referencia al sistema motor y viceversa. Casi todas las áreas dentro de la neurociencia caen dentro de estas tres

grandes divisiones: percepción, “pensamiento” y acción y poco hay en los puentes que las unen. Dada la ausencia de puentes, no queda otra para recorrer todos los territorios que andar un poco a los saltos. Esa es la historia que se encuentra aquí resumida. La de un físico, tratando de darse una idea de cual es el estado del campo de lenguaje y la comunicación, saltando entre estas tres grandes áreas y, espero, aportando algo en el camino.

1.2. Organización de esta Tesis

Este escrito esta organizado en 5 capítulos centrales (sin contar éste y un capítulo final de cierre), los capítulos que van del 2 al 6.

En el capítulo 2, estudiamos la primera de estas grandes áreas, la percepción; y en el contexto del lenguaje, la percepción del discurso. Estudiamos en particular el reconocimiento de la palabra hablada y mostramos que una ilusión auditiva conocida como el Efecto de Transformación Verbal corresponde a un caso de percepción biestable.

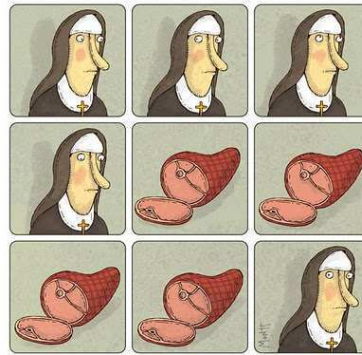
En el capítulo 3, analizamos el proceso de asociación libre como modelo para entender la transición mental entre distintas ideas. Usamos herramientas de la física estadística y modelamos el proceso como una difusión anómala en un grafo de co-ocurrencias construido a partir de un corpus de textos.

Finalmente, los capítulos 4, 5 y 6 corresponden a la última de las etapas, la de emisión. En esos capítulos, estudiamos distintos aspectos de las vocalizaciones ultrasónicas en ratas con miras a establecerlo como un posible modelo de comunicación. El capítulo 4 es un resumen de la literatura sobre vocalizaciones ultrasónicas en ratas. En el capítulo 5 estudiamos en profundidad las distintas características sonoras de las vocalizaciones y proponemos un esquema para clasificarlas automáticamente. Analizamos también como esas clases se estructuran en el tiempo y el grado de interacción vocal que hay entre distintos individuos. En el último capítulo, exploramos brevemente otros tres aspectos de la producción vocal: 1) Su relación con la locomoción, 2) Los mecanismos físicos de producción de sonido y 3) El grado de control voluntario en la emisión de vocalizaciones.

1. APERTURA

2

Biestabilidad Lexical



2.1. Introducción

La primera parte de nuestro esquema de comunicación corresponde a la recepción del mensaje y será lo que nos ocupará durante este capítulo. La pregunta central será la siguiente: ¿Cómo se pasa de la representación física de un mensaje a la representación mental? Ciertamente, el proceso será distinto según en qué forma se presente el concepto. Por ejemplo el concepto VACA, podría ser presentado con la palabra escrita (tal como se hace en esta oración) o con su secuencia de sonidos, o mostrando una imagen, etc. Nosotros estudiaremos la conversión de sonidos en palabras, es decir, la percepción auditiva de palabras. Usaremos una herramienta tradicional ya, en el campo de la neurociencia cognitiva, estudiaremos casos límite de la percepción: las ilusiones. Las ilusiones pueden parecer, a primera vista, “fallas” de nuestros sistemas perceptivos (por ejemplo, vemos moverse cosas que en realidad no se mueven, o percibimos como distintos dos estímulos que son físicamente iguales, o como iguales, cosas que son distin-

2. BIESTABILIDAD LEXICAL

tas). En realidad, lo que estas ilusiones ponen de manifiesto es lo entrenados que están los filtros de nuestra percepción para entregarnos la interpretación más plausible de acuerdo a las regularidades de nuestro ambiente. A modo de ejemplo consideremos la figura 2.1. En ella, la figura de arriba aparenta estar más atrás y ser más grande que la de abajo a pesar de que están a la misma distancia (la distancia entre los ojos del lector y esta hoja). Pareciera que nuestros ojos (o mejor dicho nuestros cerebros) nos entregan una pintura equivocada de lo que en realidad ocurre. La pregunta que nos permite entender que esa interpretación es, en realidad, óptima, es: ¿cuál es la probabilidad de que esas conjunciones de ángulos, segmentos y tonalidades ocurran en nuestras retinas, simplemente por azar en comparación con la probabilidad de que ocurran porque son producto de objetos físicos en nuestro mundo cuyas proyecciones en nuestras retinas siguen las leyes de la óptica? Vemos entonces que nosotros no recibimos de nuestros ojos la información como de una cámara de video (luminosidad de cada pixel) sino que nuestro sistema visual nos entrega una versión altamente interpretada (y mucho más útil) de los hechos. La escena visual que percibimos se compone de objetos, que pueden estar más cerca o más lejos, ser más grandes o más pequeños, etc.

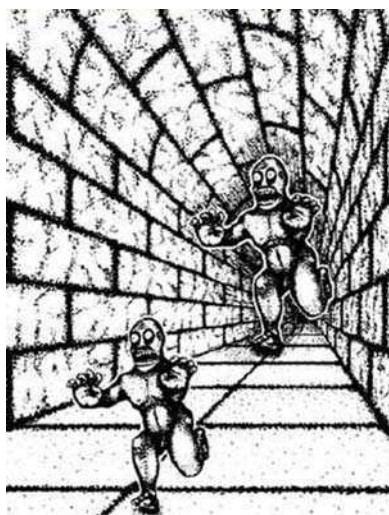


Figura 2.1: Ilusión de perspectiva - Ambas figuras tienen exactamente el mismo tamaño. La criatura más lejana aparenta ser más grande debido a la perspectiva. La misma esta sugerida por los patrones y sombras de las paredes. ²

Estudiar ilusiones nos permite, entonces, conocer la naturaleza y mecanismos de funcionamiento de nuestros filtros perceptuales. En este capítulo, analizaremos un tipo

de ilusión auditiva que se conoce como el *Efecto de Transformación Verbal* (ETV) . El equipo de Warren (1) fue el primero en describir este efecto en los años sesenta mientras estudiaba el efecto del contexto en la percepción del habla. Observaron que se produce una ilusión cuando una palabra se tiene solo a si misma como contexto. Al escuchar una misma palabra repetida muchas veces, a medida que se suceden las repeticiones, la palabra se va transformando y se comienzan a oír nuevas formas verbales. Ellos reportan, por ejemplo, que frente a la presentación repetida de la palabra *tress* (trenza en inglés) los sujetos perciben *dress, stress, Joyce, floris, florist, y purse*.

Naturalmente, surgen analogías con el sistema visual (que por cierto, está mucho más estudiado que el sistema auditivo). Los mismos Warren et al. señalan similitudes y diferencias (2) con las figuras ambiguas. Las figuras ambiguas son aquellas que tienen más de una interpretación posible. Quizá la más famosa de estas figuras sea el cubo de necker (3) (ver figura 2.2). Allí se presenta una situación bastante particular en la que es imposible decidir qué cara del cubo está por delante y cuál por detrás. El lector notará que la percepción alterna entre estas dos posibles interpretaciones; es lo que se conoce como percepción biestable.

La percepción biestable ha sido ampliamente estudiada (4) y hay resultados cuantitativos robustos sobre la frecuencia de alternancia, los tiempos de duración de cada estímulo y como éstos varían según algún parámetro que determina la plausibilidad o intensidad de cada percepto. A pesar de que la analogía con el efecto de transformación verbal fue propuesta hace varias décadas, la dificultad para manipular estímulos auditivos (en particular los verbales) hace que, hasta hoy, no haya una comparación cuantitativa que permita formalizarla.

En este capítulo propondremos una modificación sobre el ETV que permite realizar manipulaciones paramétricas con el fin de variar la estabilidad de cada percepto. Veremos así que el ETV verifica las propiedades ya establecidas para figuras ambiguas. En las siguientes secciones repasaremos algunos de los resultados ya establecidos tanto para el ETV como para la percepción biestable y analizaremos qué cosas tienen en común y cuáles no. Finalmente, discutiremos las implicaciones de estos resultados tanto para el campo de percepción del habla como para estudios más generales de la percepción, en particular, la percepción multiestable.

2. BIESTABILIDAD LEXICAL

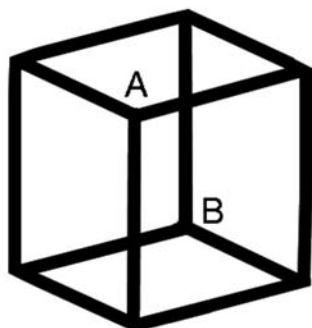


Figura 2.2: Cubo de Necker - Ejemplo de una figura ambigua. Puede interpretarse como que el vértice A está por delante del B o viceversa. La percepción alterna entre estas dos interpretaciones.

2.1.1. Sobre el Efecto de Transformación Verbal

La forma más fácil de entender el efecto de transformación verbal es, sin duda, experimentándolo. El lector puede intentarlo eligiendo una palabra al azar (como JABALÍ) y repetirla en voz alta muchas veces. A medida que transcurren las repeticiones la palabra original se desvanece lentamente para dar paso a otra interpretación de la secuencia (típicamente, con JABALI se dan VALIJA y LIJABA pero pueden ser otras más raras). En palabras de Warren et al (1):

When speech sounds – either words or short phrases– are repeated again and again without pause, then the verbal organization undergoes abrupt transition into other words and phrases, sometimes accompanied by apparent changes in the component-sounds.

Originalmente Warren et al presentaron al ETV como un análogo de las figuras ambiguas. Más tarde hicieron notar algunas de las diferencias. Por ejemplo, el ETV no se limita a patrones ambiguos, cualquier frase o palabra funciona; los cambios son imposibles de predecir; varían mucho de individuo a individuo y usualmente implican una distorsión considerable del estímulo original.

Las palabras como perceptos multiestables han recibido mucha menos atención por parte de la comunidad neurocientífica que sus contrapartes visuales, las figuras ambiguas. Esto se debe, en parte a que el campo de la audición, está, en general, menos explorado y comprendido que el de la visión. Por otro lado, usar palabras como estímu-

lo supone una dificultad adicional a la hora de intentar manipulaciones paramétricas sobre lo que se les presenta a los sujetos.

Luego de los primeros trabajos de Warren et al. (1, 2) hubo algunos intentos por estudiar las distribuciones de los tiempos pasados percibiendo cada una de las formas verbales (5). Un dato interesante es que, a pesar de que frente a la presentación de una misma palabra se pueden percibir hasta treinta o cuarenta formas verbales diferentes, la percepción queda típicamente atrapada en un par que incluye la palabra efectivamente presentada (6). Es decir, si la palabra presentada es *pez* y como formas alteradas se escuchan *paz* y *vez* la percepción oscilará mayormente solo entre *pez* y alguna de las otras dos formas: *pez, pez, pez, pez, vez, vez, vez, vez, pez, pez, vez vez vez* Esto resultará importante cuando en la sección 2.2.1 construyamos los estímulos.

La mayoría de las teorías modernas sobre el reconocimiento humano del habla (TRACE (7), SHORTLIST (8), MERGE (9), NAM (10), PARSYN (11)) presuponen que el input sonoro activa un conjunto de representaciones verbales estructuralmente similares en la memoria y que éstas compiten entre sí. Se suele definir operativamente el vecindario de una palabra a partir de la adición, sustitución o borrado de un fonema. Serían por ejemplo, vecinos de *pez*: *vez, paz, hez, pe, mes* entre otros. El ETV posibilita explorar experimentalmente en forma directa estos vecindarios. Siguiendo esta idea, Bashford y colaboradores (12) muestran que la definición de vecindario a partir de cambios en un único fonema puede ser demasiado restrictiva.

Los trabajos más recientes se centran en el estudio de los circuitos neuronales que subyacen a las alternancias del ETV. Utilizando resonancia magnética funcional, Sato et al. (13) observan, en una versión silenciosa (donde la palabra la repite el sujeto para sus adentros) del ETV, un acople funcional en el hemisferio izquierdo entre el giro frontal inferior, el giro supramarginal y el giro superior temporal sugiriendo una interacción, durante la tarea, tanto de las áreas de percepción como de producción del discurso. Kondo y Kashino (14), utilizando también resonancia magnética funcional, pero esta vez teniendo a los tiempos de las alternancias como eventos anclas, muestran una correlación positiva entre la cantidad de alternancias y la actividad en la Corteza Frontal Inferior (CFI) y una correlación negativa en el Giro Cingulado Anterior (GCA). Todo siempre lateralizado en el hemisferio izquierdo. Utilizando medidas de causalidad, establecen la siguiente secuencialidad en la percepción de las alternancias en las formas verbales: la información sensorial se envía desde las Áreas Auditivas hacia la Ínsula,

2. BIESTABILIDAD LEXICAL

allí y en la CFI se generan activamente formas verbales y luego la actividad del GCA dorsal suprime las transiciones perceptuales. Ellos concluyen que la CFI promueve la percepción de formas verbales novedosas mientras que el GCA estabiliza la percepción. Resultados similares fueron obtenidos utilizando EEG intracraneal (15).

2.1.2. Sobre la Biestabilidad Perceptual

Discutimos brevemente la noción de biestabilidad perceptual en la introducción. Enfrentados a un estímulo ambiguo, nuestros cerebros resuelven la ambigüedad proyectando en nuestra consciencia una alternancia aleatoria entre las posibles interpretaciones. A modo ilustrativo, consideremos la figura 2.3 donde se puede ver una foto de dos chitas, tomada por el fotógrafo Gerry Lemmo para National Geographic. Resulta imposible saber qué cabeza corresponde a cada chita y al observarla nuestra percepción oscila. Si se le pidiera a un sujeto que reportara qué está percibiendo en cada momento, generaría un gráfico similar al que se ve en la parte de abajo de la figura.

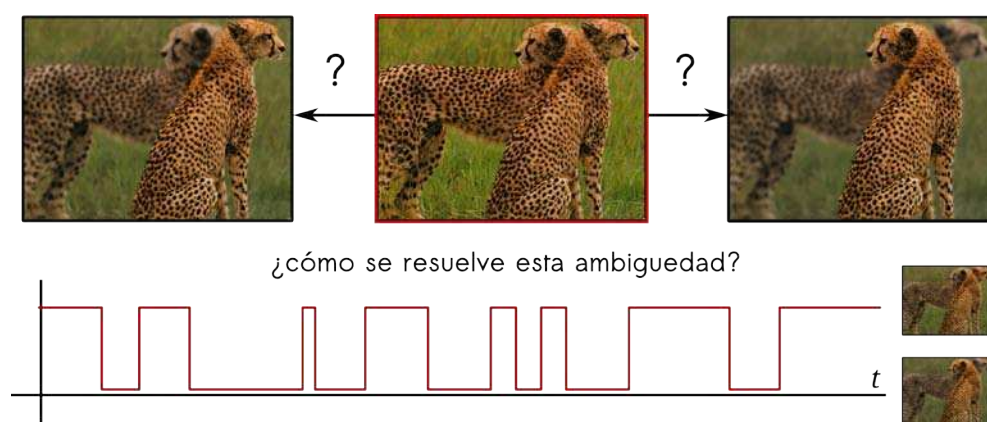


Figura 2.3: Chitas ambiguas - Otro ejemplo de una figura ambigua. Se destacan las dos posibles interpretaciones y se muestra una posible curva de respuesta de un sujeto al que se le pide que reporte momento a momento lo que percibe

La biestabilidad perceptual está mucho más estudiada que el efecto de transformación verbal, principalmente en su modalidad visual. La forma más antigua de biestabilidad perceptual descrita en la literatura es la rivalidad binocular (16). La rivalidad binocular ocurre cuando presentamos una imagen distinta (muy distinta, no como normalmente en la visión estereoscópica) a cada ojo. En lugar de ver a las dos imágenes sobreimpuestas, se ve una por un rato, luego la otra y más tarde vuelven a alternar

siguiendo un patrón temporal aleatorio. Según Wade (17), las primeras descripciones de la rivalidad binocular se deben a Giambattista della Porta quien notó que si colocaba una hoja distinta de libro delante de cada uno de sus ojos podía leer alternadamente de una página o la otra sin interferencia. Posteriormente a esto, cientos de trabajos fueron publicados estudiando la rivalidad binocular y un gran compendio y revisión de ellos se debe a Levelt (18). En ese libro, Levelt enuncia cuatro proposiciones que relacionan la intensidad con la que se ilumina cada ojo y los tiempos de alternancia.

- I Un aumento de la fuerza del estímulo en un ojo aumentará la predominancia del estímulo.
- II Un aumento de la fuerza del estímulo en un ojo no afectará l para ese ojo. (NOTA: Levelt llama l a la duración promedio que se percibe alguna de las imágenes)
- III Un aumento de la fuerza del estímulo en un ojo aumentará la frecuencia de alternancia.
- IV Un aumento de la fuerza del estímulo en ambos ojos aumentará la frecuencia de alternancia.

Donde la fuerza del estímulo es alguna medida de la saliencia del mismo; es decir, será una función creciente del contraste, decreciente de la distancia de presentación de la fovea, etc. Las proposiciones de Levelt podrían enunciarse sucintamente de la siguiente manera: *Aumentar la fuerza del estímulo en un ojo, **reduce** el tiempo medio de percepción del estímulo del ojo **contrario***. Evaluar las proposiciones de Levelt en figuras ambiguas resulta difícil porque no es evidente como definir la intensidad de cada estímulo. Un ejemplo de figura ambigua (aunque en movimiento) para la que se demostró que se cumplen las proposiciones son para el *tartán ambiguo* (ambiguous plaids) (19). Éste consiste en dos rejillas que forman un ángulo y se mueven en direcciones horizontales opuestas. Como son observadas a través de una abertura ese mismo estímulo se puede percibir como un único patrón (un tartán) que se mueve hacia arriba. La predominancia en la percepción de cada una de las interpretaciones se puede cambiar variando las velocidades y ángulos de las rejillas.

En resumen, las proposiciones de Levelt se erigen como medida estándar para mostrar que estamos frente a un fenómeno de percepción biestable. Analicémoslas, por lo

2. BIESTABILIDAD LEXICAL

tanto, un poco más en profundidad con un ejemplo. Usaremos el paradigma de rivalidad binocular que quizá torne al ejemplo más transparente. Imaginemos que estamos presentando dos imágenes distintas una a cada ojo. Llamemos 1 al ojo izquierdo, 2 al derecho y S_i al contraste lumínico de la imagen del ojo i . El espacio de estímulos será, entonces, el espacio (S_1, S_2) . Queremos ver como cambian la tasa de alternancias (TA) y la proporción de tiempo que se percibe uno de los estímulos (digamos que es el estímulo del ojo izquierdo y llamemos a esa proporción P_1) al pasearnos por el espacio (S_1, S_2) . Las proposiciones de Levelt se pueden ver más fácilmente en el espacio rotado $(S, \Delta S)$ donde $S = S_1 + S_2$ y $\Delta S = S_2 - S_1$ que corresponden a las direcciones de suma constante y diferencia constante. Leyendo las proposiciones podemos ver que si variamos ΔS manteniendo S constante, deberíamos observar un cambio en P_1 que valdría uno para valores muy negativos de ΔS y cero para valores muy positivos. Por otro lado la TA debería mantenerse constante. Por el contrario, si variamos S manteniendo ΔS constante observaríamos lo opuesto, la TA aumenta al aumentar S pero P_1 permanece constante. Esto último está ilustrado en los cuatro paneles superiores de la figura 2.4 (a, b, c, d).

Además de las proposiciones de Levelt, Logothetis (20) estableció tres características de las alternancias perceptuales observadas en todas las instancias de biestabilidad visual: exclusividad, aleatoriedad e inevitabilidad. La exclusividad establece que no se perciben ambas interpretaciones al mismo tiempo. La aleatoriedad dice que el tiempo que un estímulo permanece en consciencia es una variable aleatoria *independiente* de la duración de ese mismo estímulo en la consciencia en la percepción anterior. La inevitabilidad hace referencia a la imposibilidad de ejercer los cambios entre una percepción y su rival voluntariamente.

La exclusividad muchas veces está automáticamente garantizada, por ejemplo en el cubo de Necker, pues un vértice no puede ser a la vez cóncavo y convexo. No es así en el clásico ejemplo de la copa y las dos caras que, en principio podrían ser percibidas al mismo tiempo (aunque se podría argumentar que algo no puede ser figura y fondo al mismo tiempo). De todos modos, en el caso de las palabras queda garantizado porque los sujetos, típicamente escuchan alguna palabra, lo que descarta a las demás. Volveremos sobre la inevitabilidad en la sección de discusión. La aleatoriedad, por otro lado, es fácilmente verificable y la incluiremos entonces en nuestra figura 2.4 (e), dónde el lector podrá advertir que estamos acumulando todas las medidas cuanti-

tativas que caracterizan a la percepción biestable y que aspiramos a verificar para el efecto de transformación verbal. Notará, también, que queda una por mencionar. Se trata de la distribución de los tiempos de dominancia de cada estímulo. En todos los casos de percepción biestable reportados, la distribución de tiempos sigue o bien una distribución log-normal, o una gama (19, 21, 22). La log-normal sugiere que el proceso emerge de la multiplicación de un número grande de variables aleatorias mientras que la distribución gama surge de la concatenación de un número pequeño de procesos de Poisson. En cualquier caso, ninguna de las dos corresponde a la distribución que se obtiene de pedirle a los sujetos que aprieten un botón al azar.

Hemos resumido las características más importantes de la percepción biestable las cuales se encuentran ilustradas en la figura 2.4. Ya estamos en condiciones de pasar al experimento.

2.2. El experimento de Biestabilidad Verbal

Finalmente llegamos al experimento propuesto para mostrar que las transiciones del ETV siguen las mismas leyes que la de los estímulos biestables que discutimos en la sección anterior. El truco central reside en la construcción de los estímulos, que detallamos a continuación. Luego, explicamos el paradigma en sí y posteriormente presentamos los resultados.

2.2.1. Las ventajas del castellano en la construcción de estímulos

La parte difícil para intentar mostrar las proposiciones de Levelt suele ser la de encontrar las propiedades del estímulo que se pueden modificar paramétricamente para cambiar su estabilidad. Serían, el ángulo y la velocidad en el caso de los tartanes ambiguos. La idea para la manipulación de las palabras es la que se esconde detrás de las viñetas de Montt que ilustran el comienzo de este capítulo. Ya vimos que las transformaciones verbales tienden a formar transiciones de a pares y las oscilaciones suelen darse entre la palabra dada y la siguiente interpretación más frecuente. En castellano, hay muchas palabras bisilábicas que siguen siendo palabras válidas al invertir sus sílabas (por ejemplo, PASO y SOPA, MATE y TEMA, CASA y SACA, TACO y COTA, etc). Esto fomenta que las transformaciones se den entre estas dos alternativas y además nos da una pista de qué parámetros podemos controlar para cambiar la

2. BIESTABILIDAD LEXICAL

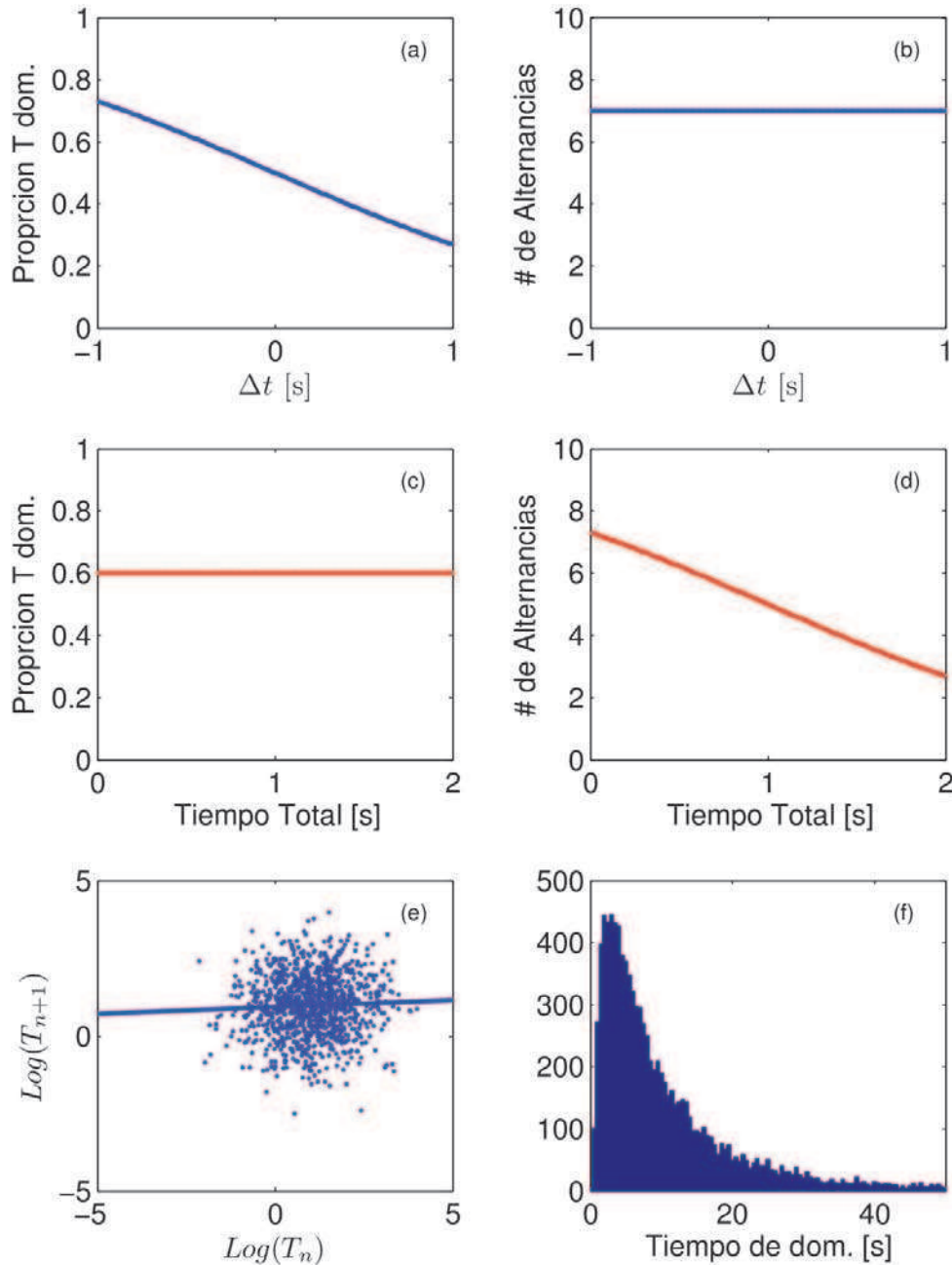


Figura 2.4: Resumen de Biestabilidad Perceptual - Acá resumimos las propiedades recolectadas de la literatura sobre la percepción biestable. De (a) a (d) corresponden a las proposiciones de Levelt. (e) ejemplifica la propiedad de aleatoriedad y (f) muestra la distribución típica de tiempos de dominancia.

2.2 El experimento de Biestabilidad Verbal

estabilidad entre ambas palabras: los tiempos entre sílabas. En la figura 2.5 vemos esto representado. Los parámetros a manipular serían los tiempos Δt_1 y Δt_2 . Para poder hacer la analogía con los S_i de la rivalidad binocular podemos hacer la analogía : $S_i \sim \frac{1}{\Delta t_i}$. Esta analogía está motivada por el hecho de que disminuir el tiempo entre MA y TE aumenta la “intensidad” o “fuerza” del percepto MATE.

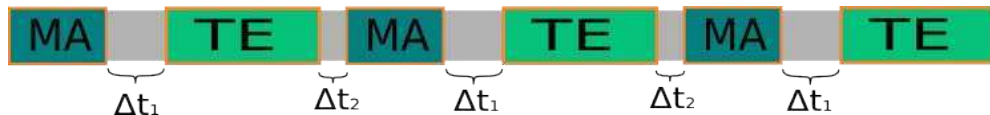


Figura 2.5: Manipulación de los estímulos - Al variar los tiempos entre sílabas se favorece una interpretación por sobre la otra.

Un tema importante a la hora de construir estímulos de habla es la naturalidad de los mismos. Esencialmente hay dos estrategias posibles: utilizar sonido sintético o utilizar grabaciones. Inicialmente nos inclinamos por el sonido sintético pero los paquetes de síntesis disponibles no eran los suficientemente flexibles para nuestros fines. Por otro lado, la grabación tampoco es la panacea, pues el cortado y pegado introduce también artefactos. Se trata del mismo problema que tienen que resolver los sistemas de anuncios automáticos en la terminal de micros o el aeropuerto. Allí, sonidos pregrabados tienen que ser empalmados para que suenen lo más natural posible. A continuación resumimos algunas de las cosas con las que hay que tener cuidado durante la construcción de los estímulos

1. Grabar todas las sílabas en una misma sesión. La voz cambia mucho con el tiempo incluso durante el transcurso de pocas horas. La voz de la mañana, no es la misma que la de la noche.
2. Grabar con referencia. Es decir, presentarle al locutor la referencia de las sílabas ya grabadas antes de grabar la nueva. Esto permite tener siempre la misma frecuencia fundamental. Lo cual se chequea posteriormente a la grabación haciendo una análisis espectral.
3. Controlar el largo de las vocales. Esto se hace con posterioridad a la grabación. Si quedaron muy largas algunas se pueden cortar con una ventana suave. La duración de la vocal es una de las claves para la acentuación y queremos tratar de que la acentuación juegue el menor rol posible.

2. BIESTABILIDAD LEXICAL

Lista de sílabas
Ba, Be, Bo, Ca, Co, Di, Fe, La, Ma, Me, Mi, Na, Ña, Ño, No, Pa, Pe, Pi, Po, Sa, Se, So, Ta, Te, Ti, Ye

Cuadro 2.1: Lista de sílabas estímulo.

Lista de palabras
casa, saca, tapa, pata, bala, lava, tana, nata, sabe, besa, pase, cepa, cabe, beca, mate, tema, cafe, feca, calle, lleca, pati, tipa, mami, mima, piña, ñapi, pizza, zapi, taco, cota, sapo, posa, paso, sopa, saco, cosa, tano, nota, baño, ñoba, peso, sope, piso, sopi

Cuadro 2.2: Lista de palabras estímulo.

Para este experimento grabamos las sílabas que se muestran en el cuadro 2.1 con una duración de vocales de 200 ± 1 ms y una frecuencia fundamental de 94 ± 5 Hz.

Con estas sílabas construimos las palabras de la tabla 2.2. El lector notará que incluimos en la lista, a modo explorativo, palabras que típicamente se usan *al vesre* en el lunfardo.

Armados con estas grabaciones, estamos listos para describir el paradigma experimental, lo cual hacemos a continuación.

2.2.2. Paradigma Experimental

El experimento consiste en la presentación de sesenta segundos de grabación con tiempos Δt_1 y Δt_2 variables. La suma $T = \Delta t_1 + \Delta t_2$ varía entre 50 ms y 450 ms y la resta $\Delta = \Delta t_1 - \Delta t_2$ varía entre -400 ms y 400 ms. Cada sujeto escucha 25 de estas grabaciones sentado frente a una computadora. Mientras escucha, se le pide que señale con la barra espaciadora qué palabra está percibiendo. A tales fines, el monitor muestra escrita alguna de las dos palabras que cambia por la otra cuando se presiona la barra espaciadora. No se les informa a los sujetos de las manipulaciones de los tiempos entre sílabas. En la figura 2.6 se resume pictóricamente el paradigma experimental.

Del experimento participaron 29 estudiantes universitarios (15 varones y 14 mujeres) con edades entre 21 y 30 años.



Figura 2.6: Esquema del paradigma experimental -

2.2.3. Resultados: El ETV es un fenómeno biestable

Dado que ya habíamos establecido en la figura 2.4 lo que esperamos de un fenómeno biestable clásico, el resultado central de esta sección será verificar cuales de las características señaladas por Levelt y Logothetis se verifican en el ETV. En la figura 2.7 se pueden ver las medidas de la figura 2.4 pero ahora para los datos del experimento.

Podemos observar (en los cuatro paneles superiores *a*, *b*, *c* y *d*) que se cumplen las proposiciones de Levelt para el número de alternancias y los tiempos medios de dominancia. En el panel (*d*) se presentan dos maneras distintas de definir las alternancias: sin normalizar (en rojo) y normalizadas por el tiempo total (en gris punteado). Esto lo hacemos para ver si lo que importa es el tiempo o la cantidad de veces que es presentada la palabra. Si solo importara la cantidad de veces que se presenta la palabra, al dividir por el tiempo total debería quedar una constante. Esto no ocurre; se pueden ver diferencias entre tiempos cortos y tiempos largos; lo que sugiere que ambos factores juegan un rol.

Los dos paneles inferiores muestran la estadística de los tiempos de percepción. En el panel *f* utilizamos un gráfico de cuantiles contra cuantiles (Q-Q plot (23)) para estudiar la similitud entre la distribución de tiempos de duración de cada percepto y una distribución normal. En este tipo de gráficos si la distribución analizada es normal los puntos deberían caer en una recta, lo cual no sucede. Sin embargo, si transformamos los tiempos a escala logarítmica la distribución se vuelve aproximadamente normal. Esto coincide con lo que se observa comúnmente en los fenómenos de percepción biestable (ver sección 2.1.2). En el mismo panel incluimos el histograma de tiempos junto con un ajuste por una curva Gaussiana y por una distribución Log-normal que puede resultar una forma de comparar más accesible a quienes no estén familiarizados con los gráficos

2. BIESTABILIDAD LEXICAL

Q-Q. Finalmente, en el panel *e* mostramos la correlación entre la duración de percepción de una de las formas verbales en dos eventos consecutivos. Observamos una correlación muy pequeña ($r = 0.12; p < 0.00001$) entre uno y otro tiempo. Este resultado es compatible con el principio de aleatoriedad de fenómenos biestables.

Todas las medidas que analizamos en la sección 2.1.2 se verifican para el Efecto de Transformación Verbal. Podemos concluir que estamos, entonces, en presencia de un fenómeno de percepción biestable. Habiendo establecido las características que el ETV comparte con otras formas de percepción biestable podemos abocarnos a estudiar sus particularidades.

¿Qué particularidades trae aparejado el hecho de que los estímulos sean palabras? Para estudiar la transición entre una forma verbal y otra podemos hacer un gráfico donde se muestre la proporción del tiempo que se escucha una de las formas verbales en función de la diferencia de tiempos ΔT . Un ejemplo para el par CASA-SACA se muestra en la figura 2.8 junto con un ajuste lineal.

Para evaluar efectos lexicales, medimos en un corpus de textos construido por nosotros (ver el capítulo 3) las frecuencias de aparición de cada palabra. Observamos una correlación positiva significativa ($r = 0.52; p < 0.01$) entre la ordenada al origen y el cociente de frecuencias en escala logarítmica entre palabras de cada par (por ejemplo entre CASA y SACA). No encontramos una correlación significativa para la pendiente. Observar una correlación positiva significa que las palabras más frecuentes son a la vez más estables. Es decir, que hace falta separar sus sílabas un tiempo mayor para que pase a dominar la palabra rival.

2.3. Discusión

Hasta hace no mucho tiempo, la percepción biestable era un fenómeno observado exclusivamente en el sistema visual. Recién en el trabajo de Pressnitzer y Hupé de 2006 (24) se documenta por primera vez la percepción biestable en la modalidad auditiva. Establecer al Efecto de Transformación Verbal como un fenómeno biestable sugiere que la percepción biestable es un principio de organización sensorial muy general; una estrategia común en nuestro sistema nervioso para manejar las ambigüedades. Un esquema en el cual, la información sensorial se acumula como evidencia de múltiples hipótesis que, en el caso de ser excluyentes, compiten entre sí y se alternan. Ahora bien, ¿Qué es

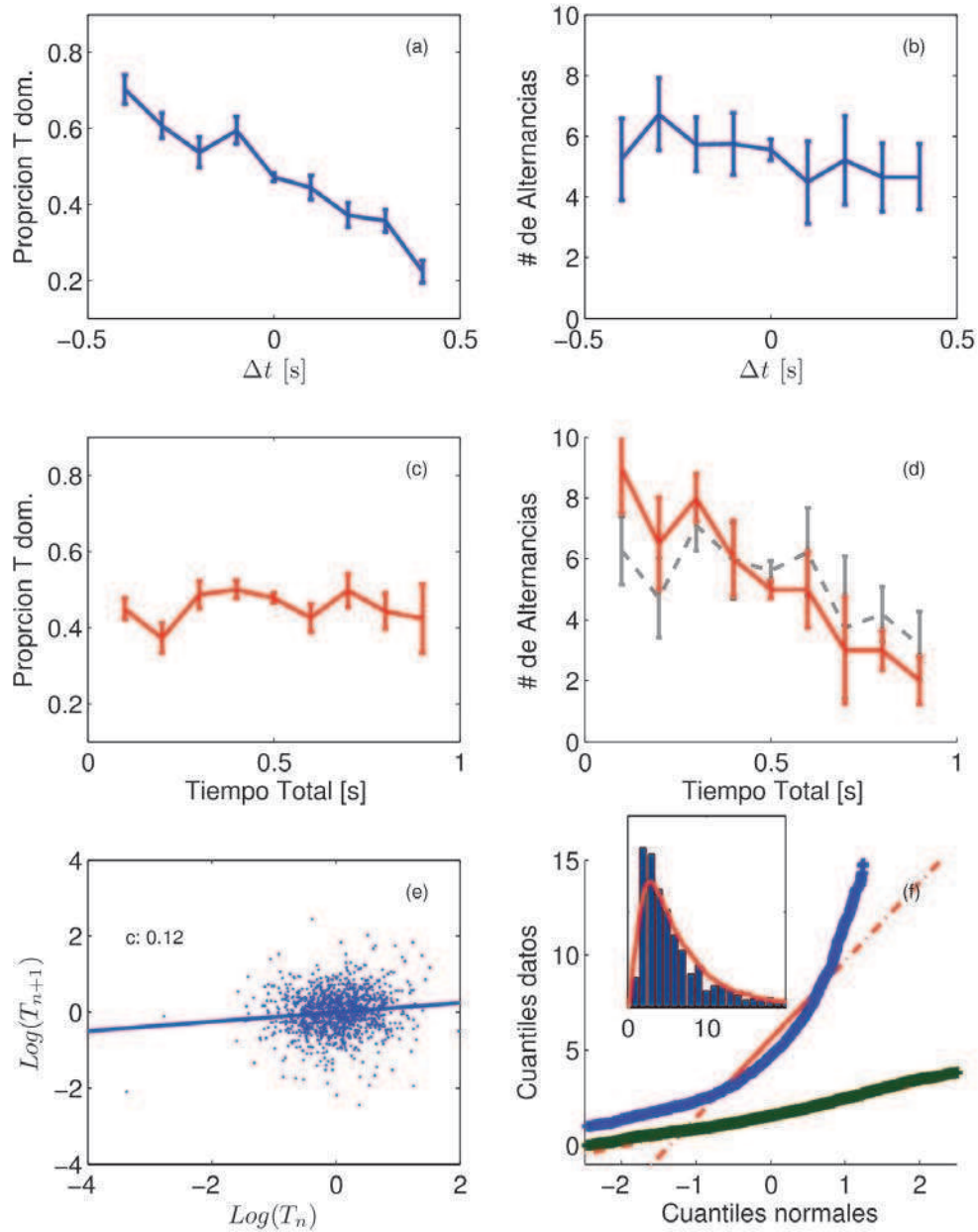


Figura 2.7: Características biestables en los datos experimentales - De (a) a (d) corresponden a las proposiciones de Levelt. En el panel (d) mostramos dos maneras de calcular las alternancias. En rojo están simplemente la cantidad de alternancias mientras que en gris punteado se encuentran normalizadas por el tiempo total. En (e) presentamos el logaritmo del tiempo de duración de un percepto en la instancia n contra la misma cantidad en la instancia $n + 1$. Finalmente, en (f) se encuentra el gráfico Q-Q para la distribución de duraciones de cada percepto (en azul) y para el logaritmo de esa cantidad (en verde). Dentro del mismo panel se presenta un histograma para esas duraciones (en segundos) junto con un ajuste con una distribución Log-Normal

2. BIESTABILIDAD LEXICAL

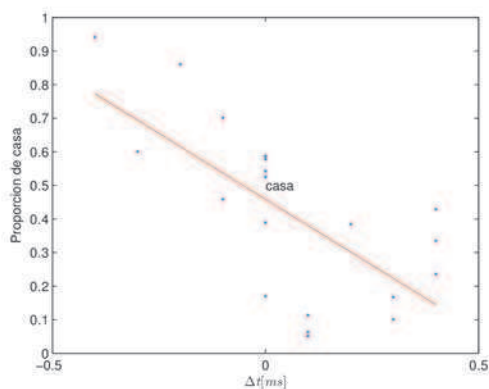


Figura 2.8: Ajuste lineal del par CASA-SACA - Ecuación de la recta $y = -0.78 X + 0.45$; $R = 0.68$

lo que alterna?

2.3.1. ¿De dónde vienen las oscilaciones?

Los modelos clásicos de percepción biestable representan cada percepto por una población de neuronas. Esas dos poblaciones de neuronas son activadas por el estímulo en un grado que depende de la intensidad de cada percepto (siempre es más fácil pensar en intensidad lumínica en cada ojo en el caso de rivalidad binocular). A su vez, ambas poblaciones se inhiben mutuamente. Solo con esos ingredientes el comportamiento sería único, y quedaría activa solamente la población que recibe el mayor *input*. Para que haya oscilaciones hay que agregar un elemento más: una adaptación lenta de cada una de las poblaciones. En ese esquema, la población ganadora comienza a adaptar lentamente (disminuye su actividad) hasta que pasa a ganar la segunda población. Ese instante corresponde a un cambio entre las dos interpretaciones del estímulo. Ahora bien, ¿Dónde están estos ingredientes en el caso de la biestabilidad lexical? Para poder responder a esta pregunta debemos discutir un poco sobre los modelos existentes para el reconocimiento de palabras.

Hay distintos modelos teóricos sobre el reconocimiento de palabras (TRACE, SHORTLIST, NAM) pero todos comparten cierta arquitectura común: son modelos con jerarquía en la escala temporal. En primera instancia se procesan las características sonoras crudas, luego los fonemas, las sílabas, las palabras y, finalmente, las frases. No todos incorporan todas las capas pero a grandes rasgos se suele distinguir entre una etapa prelexical (pre-

via a las palabras) y otra etapa lexical. La etapa prelexical recolecta evidencia sonora y activa selectivamente elementos de la etapa lexical. Los elementos de la etapa lexical compiten entre si hasta que uno supera un cierto nivel de activación y pasa a ser la palabra reconocida. La competencia se da por una inhibición mutua entre los elementos lexicales.

Bajo estos esquemas es posible entender la biestabilidad lexical de la siguiente manera. El audio repetido activa dos representaciones que compiten (por ejemplo MATE y TEMA). Si suponemos que la secuencia arrancó en MA, entonces habrá más evidencia para segmentarla como una concatenación de la palabra MATE. Eso es, en efecto, lo que los sujetos perciben inicialmente. Sin embargo, con el correr del tiempo y las presentaciones repetidas, la representación de MATE adapta lentamente (disminuye su actividad) hasta que es sobrepasada por TEMA y la percepción salta. Una consecuencia, entonces, de la mera existencia de la biestabilidad lexical es que los modelos deben incluir, además de conexiones inhibitorias a sus vecinos, alguna forma de adaptación de manera tal que presentaciones repetidas de una palabra disminuyan su activación.

En estos modelos tradicionales (TRACE, SHORT, NAM, etc), también se incorpora de alguna forma la frecuencia de uso de cada palabra de manera tal que una palabra con mayor frecuencia de uso se reconoce más rápido. Esto nos permite inferir que una palabra más frecuente será a su vez más estable y explica el efecto de correlación entre la ordenada al origen del ajuste y el cociente de frecuencias entre cada par de palabras que observamos en los datos.

2.3.2. Control Voluntario

Como vimos en la introducción, se suelen definir tres características de la percepción biestable: la exclusividad, la aleatoriedad y la inevitabilidad. Las primeras dos ya vimos que se cumplen, pero ¿qué pasa con la inevitabilidad? Originalmente la definición para rivalidad binocular decía que era imposible ejercer un control voluntario sobre las alternancias. Posteriormente se observó que la atención es capaz de modularlas; más aún, que el grado de control voluntario aumenta a medida que involucra comportamientos de más alto nivel (25). Es decir que modificar a voluntad las alternancias en rivalidad binocular es más difícil que hacerlo con las figuras ambiguas. Si bien no investigamos esto cuantitativamente es más o menos evidente, a partir de la propia experiencia, que en el caso de la biestabilidad lexical resulta fácil cambiar entre uno y otro percepto.

2. BIESTABILIDAD LEXICAL

Esto va en línea con la idea de que cuanto más alta es la función cognitiva involucrada más fácil será el control voluntario.

2.4. Conclusiones

En este capítulo estudiamos algunos aspectos del reconocimiento auditivo de palabras a través de una ilusión que se conoce como el Efecto de Transformación Verbal. Mostramos que este efecto tiene todas las características de un fenómeno de percepción biestable. También vimos que las palabras poseen una cierta estabilidad intrínseca como perceptos biestables que está dada por su frecuencia de uso. Entendimos también como los modelos de reconocimiento de habla tradicionales dan un marco para comprender de dónde surgen estas oscilaciones pero vimos que estos modelos carecen de un elemento clave: la adaptación. Esto concluye nuestra exploración de la vía de *input* en nuestro esquema de comunicación, es decir, como el sistema pasa de una serie de estímulos físicos a la representación mental de un concepto. Pero, ¿cómo se conecta una idea con otra? Ese es el material de nuestro próximo capítulo.

3

Asociación libre de palabras

3.1. Introducción

En el capítulo anterior discutimos algunos aspectos del proceso mediante el cual un concepto llega a nuestro pensamiento desde una señal externa (en ese caso una onda de sonido). En el presente capítulo analizaremos el orden de esos conceptos en nuestra cabeza. En palabras de De Saussure podríamos decir que si el capítulo anterior se ocupó de la conexión entre significante y significado, el foco de este capítulo es la organización de los significados. Como forma de explorar este espacio usaremos la técnica de asociación libre. Al presentarle a una persona una palabra y pedirle que diga la primera palabra (distinta a la presentada) que se le ocurra se observa que la palabra producida guarda, invariablemente, cierta relación con la palabra presentada. Esta técnica, junto con el estudio de redes y algunos aspectos de mecánica estadística serán las herramientas que usaremos para explorar el espacio de significados. Pero el significado ¿Qué es?

“Apart from the studies to be reported here, there have been a few, if any, systematic attempts to subject meaning to quantitative measurement. There are probably several reasons for this even in a period of intense objectivity in psychology: For one thing, the term meaning seems to connote, from most psychologists at least, something inherently nonmaterial, more akin to idea or soul than to observable stimulus and response, and therefore to be treated like the other ghosts that J.B. Watson dispelled from psychology”
Osgood - Sulci - Tannenbaum (26)

3. ASOCIACIÓN LIBRE DE PALABRAS

En su trabajo seminal Osgood et al. (26) argumentan que el espacio semántico puede ser medido cuantitativamente, determinando un cierto número de ejes principales semánticos. Cada concepto queda, entonces, representado por un vector en este espacio semántico multidimensional. Por ejemplo, es relativamente sencillo posicionar los conceptos *ratón*, *perro* y *casa* en el eje de tamaño. Si bien esta aproximación al problema lograba predecir el desempeño en varios experimentos, tenía limitaciones tanto teóricas como prácticas (27): la determinación de los ejes principales era altamente subjetiva y el número de juicios que el experimentador tenía que realizar para ubicar un término crecía proporcionalmente a la cantidad de ejes elegidos.

Ambos problemas fueron resueltos por la propuesta de construir el espacio semántico multidimensional a partir de una matriz de co-ocurrencia de términos en texto escrito (27, 28). La suposición principal detrás de esta construcción es que la proximidad semántica puede ser inferida analizando las regularidades estadísticas en un corpus de textos. Por ejemplo, si la palabra “Jirafa” aparece mencionada en un texto, es probable que las palabras, “cuello”, “cebra” y “zoológico” (que están semánticamente relacionadas) también aparezcan. Las probabilidades condicionales, se erigen entonces como una buena medida para determinar proximidad semántica. Lund y Burgess demostraron esta relación determinando que: (1) Vecinos cercanos en el espacio de co-ocurrencias corresponden a significados relacionados (2) Aglomerados (o *clusters*) en este espacio corresponden a categorías semánticas y (3) La similitud en el espacio de co-ocurrencias determina el efecto de Tiempo de Respuesta (RT) en experimentos de *priming* semántico.

Estos resultados están en línea con evidencia previa que muestra una correlación entre la frecuencia de co-ocurrencia y la probabilidad de asociación de dos términos en un experimento de asociación libre (29). La estructura asociativa de palabras juega un rol central en experimentos de memoria (30) (31) (32) (33), memoria con claves (34) y en reconocimiento de palabras (35). Más recientemente Steyvers y colegas (36) mostraron que las métricas basadas en la asociación libre de palabras tienen mejor desempeño a la hora de predecir el comportamiento humano en distintos experimentos de memoria semántica.

Si asumimos que el espacio semántico se encuentra bien caracterizado, ¿Cómo deberían ser los modelos que capturen el proceso de asociación libre a partir de la estructura de este espacio? La hipótesis subyacente en los estudios más recientes es que

la evocación de la memoria de palabras y el proceso de asociación libre son el resultado de un proceso de propagación (*concept spreading*) en el espacio semántico donde los vecinos de este espacio multidimensional son algo análogo a un campo semántico (37, 38, 39, 40, 41, 42, 43).

Bajo este esquema, significados que no están directamente conectados, pueden relacionarse a través de cadenas largas de eslabones semánticos (por ejemplo, la secuencia *leon - felino - tigre - rayas*). Estas trayectorias semánticas pueden ser reconstruidas a partir de asociaciones de a pares como el camino más corto que conecta dos significados que no están inmediatamente relacionados (44, 45). Trabajos previos muestran que las asociaciones indirectas juegan un rol en la memoria inducida por claves (45) y el reconocimiento de palabras (35).

Esta evidencia es consistente con la idea de que el proceso de asociación de palabras y la evocación de memorias involucran la navegación en un grafo semántico. En este capítulo nos proponemos estudiar justamente eso: estudiar cómo es que cadenas largas de asociaciones se embeben en el grafo semántico y modelar el proceso de asociación como un proceso difusivo (veremos más adelante que será un proceso difusivo anómalo).

En las secciones que siguen, describimos la construcción de la métrica para definir el grafo semántico y el experimento de asociación libre. Finalmente, analizamos distintos modelos de difusión posibles.

3.2. Una métrica para el espacio de palabras

Habiendo establecido que vamos a usar el grafo de co-ocurrencias como sustrato para estudiar el proceso de asociación libre, falta decidir qué textos usaremos y definir más precisamente la noción de co-ocurrencia. En los siguientes apartados describimos la construcción del corpus y formalizamos la definición de la métrica.

3.2.1. Construcción del Corpus

Para la construcción del corpus necesitamos grandes volúmenes de textos en castellano. Internet nos provee de una fuente casi ilimitada. Incluimos en nuestra base de datos: la wikipedia completa en castellano, todos los artículos del diario La Nación entre 2000 y 2011, todos los libros en castellano que no se encuentran protegidos por derechos de autor del Proyecto Gutenberg. Todos los textos fueron “limpiados” de

3. ASOCIACIÓN LIBRE DE PALABRAS

código HTML y también se les quitaron las tildes. Esto fue hecho para captar errores ortográficos comunes ya que las tildes suelen omitirse con frecuencia especialmente en Internet. Por supuesto, el costo de esta operación es que se pierde la distinción de palabras que difieren solo en su acentuación (por ejemplo te y té). El corpus contiene alrededor de mil millones de palabras.

3.2.2. Definición de la métrica entre conceptos

Como mencionamos en la introducción, para definir la distancia entre términos usaremos la probabilidad condicional; es decir, la probabilidad de encontrar en un texto a la palabra A dado que sabemos que en ese texto está presente la palabra B. La distancia entre A y B estará dada entonces por:

$$D(A \rightarrow B) = \frac{1}{P(B|A)} \quad (3.1)$$

Una cosa importante a notar es que la probabilidad condicional no es simétrica ($P(A|B) \neq P(B|A)$) y por lo tanto D no es formalmente una distancia. Podemos estimar la probabilidad condicional a partir del conteo de los eventos de co-ocurrencia de la siguiente manera:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \sim \frac{N(A \cap B)}{N(B)} \quad (3.2)$$

Falta aclarar que consideramos una co-ocurrencia. Diremos que dos palabras co-ocurren si se encuentran a una distancia menor o igual a diez palabras entre si.

Todos los algoritmos para la construcción del corpus, limpiado de textos y conteo de co-ocurrencias fueron escritos en PERL y están disponibles en la página web del laboratorio (neurociencia.df.uba.ar).

3.3. Experimentos de asociación libre

El experimento de asociación libre es conceptualmente sencillo. La idea es que cada sujeto recibe una palabra y responde la primera palabra que se le ocurre. Para generar una cadena cada palabra que produce un sujeto se usa como palabra estímulo de otro. De esta forma se va generando una trayectoria de asociaciones como en una especie de juego de cadáver exquisito pero de una sola palabra. Dado que la dinámica del

experimento es razonablemente lúdica, decidimos implementar el experimento como un juego *online*. Esto nos permitió obtener muchos datos de varios participantes en un tiempo relativamente corto.

3.3.1. Plataforma para experimentación Online

El juego de asociaciones fue programado usando una combinación de HTML, PHP y Javascript. Los datos fueron guardados en una base MySQL. El sitio estaba dividido en tres grandes partes solo dos de las cuales eran accesibles para los jugadores. La primera parte consistía en una página de registro para jugadores nuevos. Allí los usuarios deben dejar su nombre y su dirección de correo electrónico donde reciben la confirmación de su registro y las instrucciones del juego.

La segunda y principal parte del sitio era la página personal de cada jugador. Allí podían ver cuantas palabras habían recibido y enviar sus asociaciones a otros jugadores. Por ejemplo si un jugador recibía la palabra PERRO, asociaba libremente (GATO por ejemplo) y enviaba su palabra a un nuevo jugador. Esto iba formando, con cada paso, una trayectoria de asociaciones. Cada tanto, los jugadores tenían la posibilidad de “doblar” y enviar la palabra asociada a dos jugadores, bifurcando así la trayectoria. Esto implica que algunas asociaciones pertenecen a más de una trayectoria. La probabilidad de bifurcación podía ser controlada dinámicamente para regular el tráfico total de palabras en el juego.

Luego de registrarse, los jugadores recibían una palabra “semilla” de una lista cerrada de veinte palabras (ver tabla 3.1). Una vez que un jugador hubiese contestado todas sus palabras pendientes podía robar palabras de otro jugador o ver una instantánea de su trayectoria. Todo esto fue hecho para darle una dinámica de juego e incentivar la participación. Si los jugadores dejaban muchas palabras sin responder recibían primero una tarjeta amarilla y posteriormente una roja (lo que implicaba eliminarlos del juego). Cada 3 días se enviaba un recordatorio automático indicando a cada jugador cuantas palabras pendientes tenía.

Una tercera parte del sitio que no era accesible a los jugadores servía de “puerta trasera” del sitio para monitorear el desarrollo del juego. Desde ahí podíamos: ver la actividad de cada jugador, del juego como un todo; enviar correos de recordatorio; sacar tarjetas amarillas y rojas y cambiar la probabilidad de bifurcación de las trayectorias.

El juego estuvo abierto por 2 meses y generó 11.000 asociaciones de 120 jugadores.

3. ASOCIACIÓN LIBRE DE PALABRAS

Lista de palabras semilla
bebe, anillo, torta, almohadón, agua, motor, escoba, huevo mano, hoja, planta, caramelo, juez, mapa, barco, sonido, murciélago rosas, lengua, pluma, menta, chocolate, casa, rodilla

Cuadro 3.1: Lista de palabras semilla en el juego de asociación.

3.4. Resultados

Quizá una de las cosas más difíciles con datos altamente multidimensionales sea elegir una forma de visualizarlos que resulte clara. El espacio semántico que definimos en la sección anterior tiene muchas dimensiones, ¿Cómo representar las trayectorias de asociaciones que obtuvimos en el juego para intentar visualizar su estructura?

3.4.1. Fuzzy clustering y proyección de Sammon

Hay muchas técnicas de reducción de la dimensión que permiten “aplanar” los espacios. Nosotros elegimos hacer una primera reducción encontrando grumos o *clusters* de palabras muy relacionadas entre sí, lo que intuitivamente llamamos categorías de palabras. Ahora bien, una palabra suele poder pertenecer a más de una categoría y por lo tanto elegimos usar algoritmos de *clustering* borrosos; es decir, que le asignan a cada palabra un grado de membresía a cada *cluster* o grupo que vale 0 si la palabra no pertenece a ese grupo y que crece con el grado de pertenencia hasta llegar a un máximo de 1.

Siguiendo a Hotta et al (46)) aplicamos un método espectral (de descomposición en autovectores) como una forma de encontrar los vectores de membresía. El método propuesto por ellos es jerárquico, es decir que extrae los *clusters* de mayor a menor grado de cohesión. Graficar la cohesión del *cluster* i -ésimo en función del orden en que fue extraído permite definir un criterio para decidir cuantos *clusters* significativos hay en el grafo.

El *clustering* borroso permite hacer una primera reducción de la dimensionalidad. Si tenemos un grupo de N palabras, la matriz de co-ocurrencias definirá un espacio N -dimensional donde se ubica cada una. Usando el *clustering* borroso podemos pasar de un espacio de dimensión N a otro de dimensión K (donde K será el número de *clusters* significativos, con K mucho más chico que N). Supongamos que encontramos 5 *clusters* significativos. Cada palabra será, entonces, un punto en un espacio 5-dimensional donde

la coordenada i -ésima nos dice el grado de membresía al *cluster* i -ésimo.

Para poder llevar este espacio a algo fácilmente representable (2 dimensiones) usamos un *mapeo* no lineal conocido como Proyección o Mapa de Sammon (47). Esta proyección consiste en encontrar la representación 2D que minimiza la deformación de las distancias entre los datos.

¿Cómo se verá, entonces, una trayectoria representada en dos dimensiones? La respuesta está en la figura 3.1. Ahí se puede ver una trayectoria del juego dibujada en el espacio 2D reducido. La trayectoria empieza en “huevo” (marcada con una estrella) y termina en “Sweater” (marcada con un cuadrado). Algunos aspectos de la topología de la trayectoria, que más adelante cuantificaremos, resultan evidentes.

Primero, podemos observar que la mayoría de las transiciones en la trayectoria resultan en un salto pequeño (es decir, una asociación entre dos palabras cercanas). Segundo, para varias asociaciones consecutivas, la trayectoria queda confinada en un “barrio” (una categoría) del espacio de palabras. Tercero, dentro de cada barrio, los ciclos son muy prevalentes. En efecto, los más frecuentes son ciclos de orden 2, es decir dos palabras con probabilidades de regreso recíprocamente altas (por ejemplo: pelota, fútbol, pelota ... u mujer, hombre, mujer...). Finalmente, en algunas instancias esporádicas, una asociación resulta en un salto muy largo que relaciona palabras de distintos barrios.

3.4.2. Los ciclos y el comportamiento difusivo

Para cuantificar lo observado en la sección anterior definimos, a partir de la estadística del corpus, el desplazamiento entre la palabra k y la palabra j como:

$$\Delta(k, j) = \frac{1}{A P(j|k)} \quad (3.3)$$

con

$$A = \sum_{\forall k, j} \frac{1}{P(j|k)} \quad (3.4)$$

Se suele referir a esta medida como distancia, aunque, como vimos, no es formalmente una distancia pues no es simétrica.

Analizaremos el ensamble de trayectorias $\{T_i\}$ que no es más que una lista de

secuencias. T_i es una secuencia de palabras $\{T_{i_1}, T_{i_2}, \dots, T_{i_{L_i}}\}$ donde L_i es la longitud de la trayectoria T_i . Definimos entonces $\langle \Delta(n) \rangle$ como el promedio de $\Delta(T_{i_j}, T_{i_{j+n}})$ (es decir el desplazamiento entre la palabra T_{i_j} y la palabra $T_{i_{j+n}}$) sobre todas las trayectorias posibles.

La segunda medida que vamos a usar es la probabilidad de regreso a una dada palabra, que se puede definir sencillamente como la fracción de trayectorias que regresan a la palabra original después de n asociaciones.

En la figura 3.2 presentamos los desplazamientos medios y fracción de ciclos en función de la cantidad de pasos (o asociaciones) intermedias.

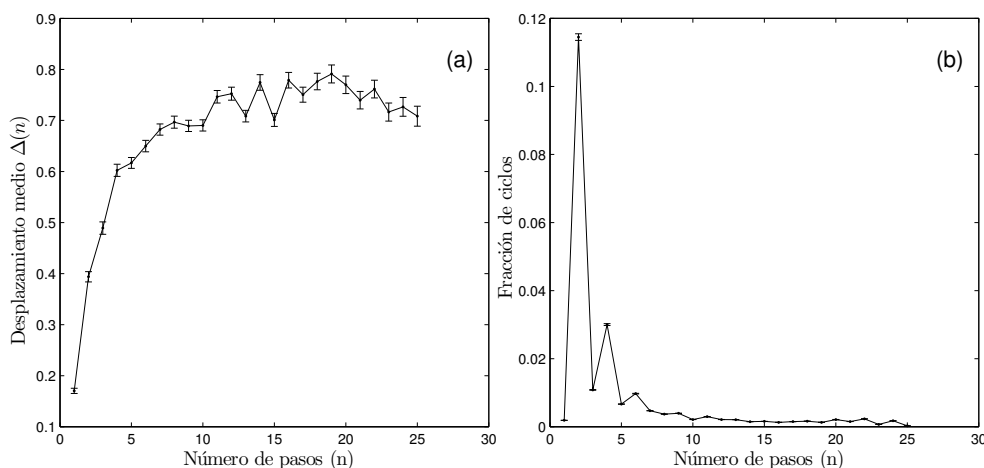


Figura 3.2: Datos Experimentales - (a) Desplazamiento medio para cada paso calculado según la ecuación 3.3 (b) Probabilidad de regreso a la palabra original después de n pasos.

Como se puede ver, el desplazamiento aumenta monótonicamente con el número de pasos n , lo que refleja la pérdida de memoria luego de una concatenación de asociaciones libres. Esta progresión alcanza un valor asintótico de aproximadamente 0.8, un valor menor a 1 que indica que dos palabras cualesquiera dentro de una trayectoria están en promedio más cerca que dos palabras al azar dentro del grafo. Es decir que las trayectorias no recorren el grafo de manera uniforme, dicho de otra forma, el sistema no es ergódico.

La fracción de ciclos también presenta un patrón claro. Hay un efecto de paridad marcado (números pares de pasos tienen una probabilidad de retorno más alta) modulado por un decaimiento exponencial. Este efecto de paridad indica que la estructura

3. ASOCIACIÓN LIBRE DE PALABRAS

de ciclos está dominada por ciclos de orden 2, es decir que segmentos de la forma perro-gato-perro... son muy frecuentes en trayectorias de asociación libre. Cabe destacar que la probabilidad de encontrar un ciclo de orden 2 es relativamente alta (alrededor del 12 %, indicando que aproximadamente 1 en 8 de las asociaciones regresan a la palabra original). El hecho de que no se observen ciclos de orden 1 simplemente muestra que los sujetos cumplen con la instrucción de no repetir la palabra que reciben.

Estas curvas que caracterizan el desplazamiento y la estructura de ciclos nos servirán de vara de medida para evaluar modelos generativos de asociación libre. Describiremos estos modelos en términos de cadenas de Markov donde cada palabra representa un estado y el modelo queda determinado al especificar las probabilidades de transición entre estados.

En la siguiente sección analizaremos distintas propuestas para las probabilidades de transición y simularemos trayectorias para contrastar las curvas de desplazamiento y ciclos simuladas contra los datos experimentales.

3.5. Modelos difusivos para la asociación libre

Modelaremos el proceso de asociación como una cadena de Markov de orden uno, es decir que sólo tiene memoria del estado actual del sistema. Una cadena de Markov de orden uno es un proceso aleatorio en el que un sistema hace una transición entre diferentes estados y la probabilidad de saltar a un dado estado depende únicamente del estado en que nos encontremos. En nuestro caso, cada palabra sería un estado y la transición entre estados corresponde al proceso de asociación.

¿Cómo elegir las probabilidades de transición entre palabras? Supondremos que serán alguna función de la distancia entre conceptos. Así, debería ser más probable saltar de “perro” a “gato” que a “electrón”. En los siguientes apartados estudiaremos esencialmente dos tipos de formas funcionales. En la primera parte, analizaremos funciones de vecindario acotado, es decir donde cada palabra tendrá un número definido de vecinos. En la segunda, estudiaremos dependencias funcionales libres de escala (leyes de potencias).

3.5.1. Modelos de vecindario acotado

Quizá el modelo más sencillo que se nos podría ocurrir corresponde a un modelo de núcleo en el que la probabilidad de transición desde una palabra dada es una función escalón. Supongamos que $\{1\dots N\}$ es la lista ordenada por distancia de los vecinos de una palabra. Entonces, la distribución de probabilidad en este conjunto para un modelo de núcleo de K vecinos, está dada por:

$$(3.5) \quad P(i) = \begin{cases} \frac{1}{K} & i \leq K \\ 0 & i > K \end{cases}$$

donde $P(i)$ es la probabilidad de saltar al i -ésimo vecino. Los comportamientos límites de este modelo son fáciles de entender. $K = 1$ corresponde a un comportamiento de perseverancia, es decir, asociar siempre la palabra más cercana en el grafo. Siempre que dos palabras sean recíprocamente sus vecinas más cercanas (lo cual ocurre seguido pero no siempre pues el grafo no es simétrico) la trayectoria queda atrapada en un ciclo. Por lo tanto, en este límite, los ciclos son muy prevalentes y el desplazamiento no converge al promedio. El otro caso límite corresponde a $K = N$ (el número total de nodos en el grafo) en el cual las transiciones se hacen completamente al azar. En esta situación, los ciclos son muy raros (orden $\frac{1}{N}$) y el desplazamiento es chato e igual al desplazamiento medio.

Usando una medida de cuadrados de diferencias entre las simulaciones y los datos exploramos qué valores de K son compatibles con lo observado para los ciclos y el desplazamiento. Un resumen de esta exploración puede verse en la figura 3.3 en los paneles a, b, c, d, e y f en color azul.

Medimos la bondad del ajuste usando errores cuadráticos normalizados definidos como: $\bar{e}_i = \frac{e_i}{\mu_i}$, donde e_i es el error cuadrático para la curva de desplazamiento ($i=1$) y la curva de ciclos ($i=2$) respectivamente. Los $\mu_i = \sqrt{x_{i1}^2 + x_{i2}^2 + \dots + x_{in}^2}$ son factores de normalización (x_{ij} son puntos de la curva i).

Las curvas de ciclos y desplazamiento pueden ser ajustadas razonablemente de manera individual pero con valores drásticamente distintos del parámetro K . El mejor

3. ASOCIACIÓN LIBRE DE PALABRAS

ajuste para los ciclos corresponde a $K = 8$ ($\bar{e}_1 = 0.66$ y $\bar{e}_2 = 0.11$) mientras que el mejor ajuste para la curva de desplazamiento se da para un valor de $K = 71$ ($\bar{e}_1 = 0.05$ y $\bar{e}_2 = 0.68$). No nos fue posible encontrar valores de K intermedios que permitieran ajustar ambas curvas simultáneamente.

Esto quizá se deba a que pesamos igualmente a todos los primeros K vecinos sin importar cuan cerca o lejos estén. Para explorar esta posibilidad, decidimos simular un modelo en donde la probabilidad de transición dependa exponencialmente de la distancia. En este modelo, las probabilidades de transición entre la palabra k y la j estarán dados por:

$$P(k \rightarrow j) = Ae^{-\frac{\Delta(k,j)}{T}} \quad (3.6)$$

donde A es una constante de normalización y T puede ser pensada como una “temperatura” del sistema, es decir que representa el grado de estocasticidad. En este modelo T juega un rol similar al de K , con las analogías $K = 1$ ($T \rightarrow 0$) para un comportamiento de perseverancia y $K = N$ ($T \rightarrow \infty$) para saltos completamente aleatorios. Sus casos límites son, de hecho, idénticos.

Al igual que para el modelo anterior, encontramos dos valores del parámetro T que podían explicar razonablemente bien la curva de ciclos o de desplazamiento por separado ($\bar{T} = 7.4$; $\bar{T} = 66.6$, donde la barra denota que T está normalizada por el Δ medio del grafo). Sin embargo, como observamos para el modelo de núcleo, no nos fue posible encontrar un valor intermedio de T que permita ajustar ambas curvas simultáneamente. Los valores de \bar{e}_1 y \bar{e}_2 están graficados en la figura 3.3 en el panel (g). De ese gráfico, resulta evidente que los mínimos de ambas curvas no se solapan y por lo tanto no existe un valor de T que garantice un buen ajuste para las dos curvas. Un intento de hacer el gráfico para un valor intermedio (el que minimiza $emax = \max\{\bar{e}_1, \bar{e}_2\}$) tampoco resulta satisfactorio (ver paneles (b) y (e)). Obtuvimos un resultado muy similar para el modelo de núcleo.

Si bien, no exploramos esto exhaustivamente, es cualitativamente fácil de entender que cualquier modelo en el que las probabilidades de asociación estén confinadas a un núcleo o vecindario finito (aún con pesos distintos) no puede explicar simultáneamente las curvas de ciclos y desplazamiento.

La probabilidad de ciclos de orden 2 es igual a $\frac{1}{K}$ (si cada palabra puede asociarse a

K vecinos, entonces la probabilidad de regresar a la palabra original es $\frac{1}{K}$), esto impone un valor fijo y pequeño para K . Por otro lado, la curva de desplazamiento (que muestra una rápida convergencia a un valor asintótico) determina que el número de vecinos (K en el modelo de núcleo o vecinos efectivos en el modelo exponencial) debe ser grande ~ 70 . Por lo tanto, la asociación de palabras no puede ser explicada por un modelo difusivo común acotado. Esto va en línea con las observaciones cualitativas que hicimos en la sección anterior donde observamos que si bien la mayoría de las asociaciones están restringidas a un entorno local en un *cluster* semántico, algunas pocas asociaciones son de largo alcance y relacionan palabras de distintos barrios

Una forma posible de conseguir este comportamiento sería agregarle a los modelos anteriores una pequeña probabilidad de producir, cada tanto, saltos completamente aleatorios. Esta metodología sirve, es decir, es posible ajustar ambas curvas con modelos de este estilo pero viene con el costo adicional de tener que ajustar un parámetro extra (la probabilidad de saltos aleatorios).

Una mejor solución es considerar procesos difusivos anómalos en los que la distribución de probabilidad es invariante de escala y está dada por una ley de potencias. Este caso, es el que estudiaremos en la siguiente sección.

3.5.2. Modelos libres de escala

Muchos fenómenos naturales siguen distribuciones donde la probabilidad de eventos de gran magnitud es mucho mayor de lo que esperaríamos dada una distribución normal (Gaussiana). Existe una familia de procesos aleatorios que reflejan matemáticamente este hecho: los *Lévy flights*. Los *Lévy flights* (48) son procesos de caminata al azar en los que la distribución de pasos sigue una ley de potencias (del tipo $x^{-\alpha}$). Se las llama distribuciones libres de escala pues al cambiar x por Ax su forma no cambia. Son distribuciones con “colas largas”, es decir que decaen lentamente y por lo tanto saltos de gran magnitud son posibles. A modo de comparación con los modelos anteriores presentamos en la figura 3.4 las densidades de probabilidad de los modelos de núcleo, exponencial, y libre de escala.

Si repetimos las simulaciones y buscamos el parámetro α que mejor ajusta la curva de desplazamientos y la de ciclos vemos que ahora coinciden para un dado valor ($\alpha = 0.9 \pm 0.1$). Esto es lo que presentamos en la figura 3.5. Vemos que es posible, entonces, encontrar un único valor del parámetro que ajusta razonablemente bien ambas curvas.

3. ASOCIACIÓN LIBRE DE PALABRAS

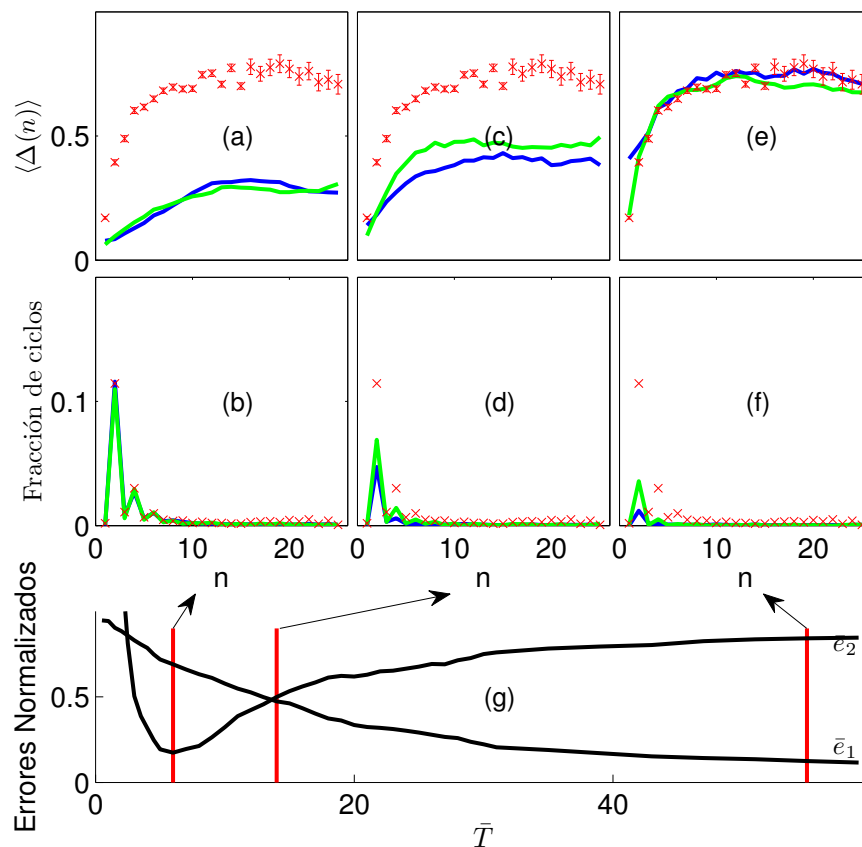


Figura 3.3: Simulaciones para vecindarios acotados - En los paneles superiores (a-f) mostramos las curvas de desplazamiento y ciclos junto con los ajustes para los dos modelos de vecindario acotado. En verde, modelo de K vecinos; en azul, modelo exponencial. En el panel inferior (g) se muestra la curva de errores normalizados para el modelo exponencial y se indican los valores que corresponden a los ajustes de los paneles superiores.

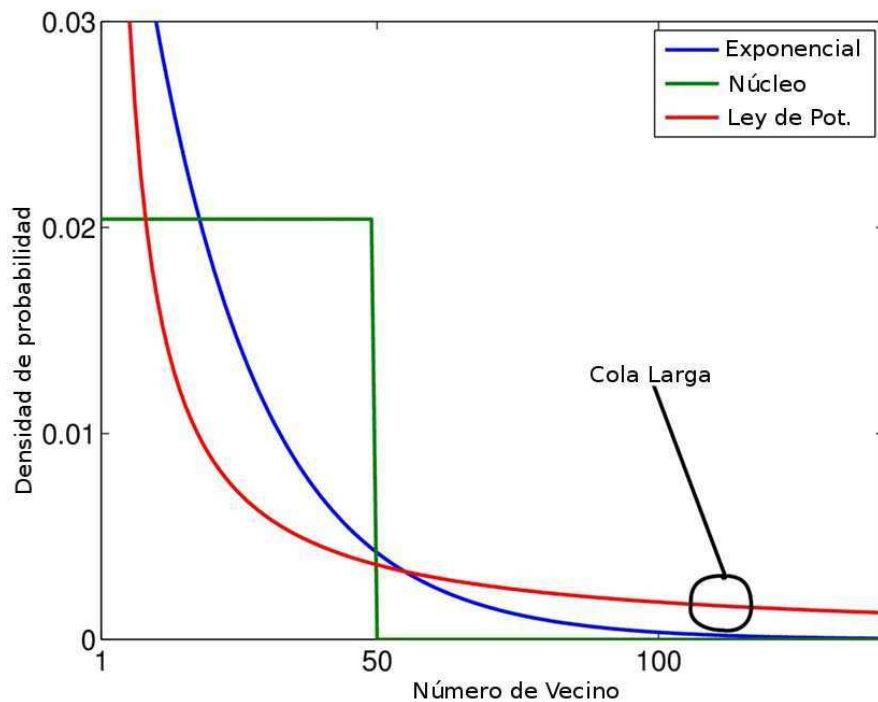


Figura 3.4: Densidades de probabilidad de los distintos modelos - Se puede ver como un modelo de ley de potencia decae mucho más lentamente que los otros dos.

Antes de cerrar esta sección nos gustaría hacer un pequeño comentario de algo que quizá ya hayan notado. Existe otra forma de interpretar estos datos, sin pasar por la noción de distancia. En efecto, si definimos nuestra pseudo-distancia $d_{i,j}$ como $d_{i,j} = \frac{1}{p_{i,j}}$, al decir que la probabilidad de transición es una ley de potencias de exponente negativo estamos, en parte, invirtiendo esta relación. Es decir que el proceso está dado por las probabilidades $p_{i,j}^\alpha$. Dado que alfa es menor a 1, esto quiere decir que la distribución está más “achatada” (de nuevo sirve pensar en los límites cuando $\alpha \rightarrow 0$ y $\alpha \rightarrow \infty$). Bajo esta interpretación, el proceso de asociación está gobernado por las mismas probabilidades de transición que el proceso de escritura pero escaladas de manera tal de alargar la cola de la distribución. Una nueva inspección a la figura 3.1 puede darnos una pista sobre el origen de las colas largas. Las transiciones largas entre barrios parecen ser frecuentes en palabras polisémicas (palabras con más de un significado posible). En el ejemplo de la figura, *boca* puede referirse a la parte de la cara o al equipo de fútbol. Durante el discurso coherente evitamos la ambigüedad que puedan traer estas palabras manteniéndonos en un único contexto; a menos que queramos hacer un chiste o juego

3. ASOCIACIÓN LIBRE DE PALABRAS

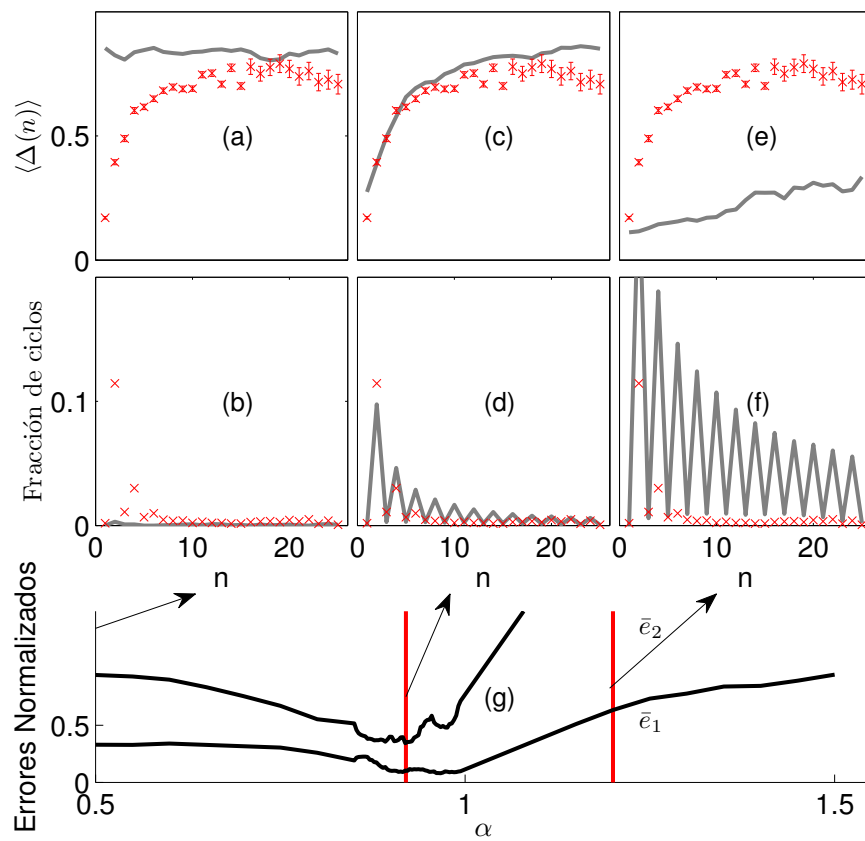


Figura 3.5: Simulaciones libres de escala - A diferencia de los modelos de vecindario acotado, los modelos libres de escala tienen una región de su parámetro α que ajusta tanto a la curva de desplazamientos como a la de ciclos.

de palabras. Como ese capitán que pidió que bajen las velas y todos los que estaban en la parte de arriba del barco quedaron a oscuras; o como en la novela de Saer, donde no se sabe si por falta de habilidad o por miedo al agua pero nadie nada nunca. En cambio, el proceso de asociación libre nos permite esos saltos entre barrios que parecerían favorecidos por la existencia de palabras polisémicas. Este *ansatz* va en línea con trabajos que muestran que la relación de polisemia convierte a la red semántica en una red de mundo pequeño??.

3.6. Conclusiones

En este capítulo analizamos trayectorias de asociación libre, obtenidas a través de un juego online, embebiéndolas en un grafo de co-ocurrencias derivado de un Corpus grande de textos. Usamos esta estrategia para comparar la producción lingüística altamente estructurada de un texto con las trayectorias más azarosas producidas mediante la asociación libre. Pudimos observar que estas trayectorias no pueden ser explicadas por un modelo difusivo acotado, ya que la estructura de ciclos y de desplazamientos no pueden ser ajustadas para un mismo valor del parámetro que determina el tamaño del vecindario. Sin embargo, esto puede ser solucionado usando distribuciones libres de escala. Esto es consistente con nuestras observaciones cualitativas del proceso de asociación libre: las asociaciones, en su mayoría, están confinadas a pequeños *clusters* semánticos y por lo tanto son estereotipadas y los ciclos son prominentes; en algunas instancias se producen saltos más largos que unen distintos *clusters* y que posiblemente sean más frecuentes en palabras polisémicas.

Con esto cerramos los estudios hechos en humanos y pasamos a la tercera y última parte de esta tesis que corresponde a la emisión de un mensaje.

3. ASOCIACIÓN LIBRE DE PALABRAS

4

Vocalizaciones Ultrasónicas en Ratas.

En los capítulos anteriores estuvimos discutiendo algunos aspectos de la decodificación de sonidos en palabras y de las transiciones mentales entre palabras o conceptos. En los capítulos que siguen nos enfocaremos en la tercera pata de nuestro esquema de *recepción - elaboración - emisión* para la comunicación. Nos abocaremos, entonces, a la producción de sonidos con fines comunicacionales pero ya no en la especie humana. Estudiaremos la producción vocal de las ratas en el rango ultrasónico. Supongo que esto parecerá un poco extraño y de hecho lo es. Este salto corresponde a un interés personal, una necesidad de estudiar un sistema más sencillo (o al menos eso pensé en ese momento) y donde se puede acceder más directamente a los correlatos fisiológicos. Hay también, como en casi toda la vida, una cuota de aleatoriedad y ocurre que es una de las líneas que comenzaba cuando llegue al laboratorio de Diego Laplagne. El campo de las vocalizaciones ultrasónicas está relativamente subexplorado y, dado el uso extensivo de ratas como modelo animal de laboratorio, resulta interesante conocer los detalles de funcionamiento y complejidad de ese sistema de comunicación que no podemos escuchar. Ahora sí, sin más, podemos hacer un breve repaso del estado actual del campo.

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

4.1. Introducción

Las vocalizaciones ultrasónicas en la rata fueron descritas por primera vez por John Anderson en 1954 (49). Los primeros registros eran bastante crudos ya que no se disponía de equipos comerciales capaces de grabar y almacenar audio por encima de los 20 KHz. Era común utilizar “detectores de murciélagos”, dispositivos que permiten dividir la frecuencia de la vocalización para llevarla al rango audible. Con el abaratamiento de los costos y la mejora de los equipos de grabación, el campo creció rápidamente. En seguida se apreció el fuerte componente emocional que tiene la emisión de vocalizaciones observando variaciones en la producción en distintos contextos: agresión, juego, apareamiento, etc. También se comenzó a estudiar los mecanismos físicos de producción que, aún hoy, no están del todo comprendidos. Posteriormente aparecieron estudios sobre los circuitos neurales involucrados en su producción. Estos son algunos de los temas que discutiremos en este capítulo.

En las secciones siguientes analizaremos algunas de las características sonoras de las vocalizaciones y veremos hasta qué punto éstas nos permiten definir categorías. Repasaremos también las observaciones que arrojan luz sobre las posibles funciones sociales de las vocalizaciones. Discutiremos además, qué se sabe hasta ahora de los circuitos neuronales involucrados en su producción y, finalmente, qué grado de control voluntario poseen sobre su aparato fonador.

En la última sección de este capítulo describimos, también, el dispositivo experimental que se utilizó para grabar la mayoría de las vocalizaciones analizadas en esta tesis.

4.2. Características sonoras

4.2.1. Rangos, duraciones y propiedades espectrales

Las ratas adultas emiten vocalizaciones ultrasónicas en el rango que va desde los 20 KHz hasta los 80 KHz aproximadamente. Sus duraciones son muy variables y se encuentran en el rango que va de los 20 ms hasta los 2 segundos. Presentamos espectrogramas de algunas vocalizaciones ultrasónicas en la figura 4.1. Una característica que se observa inmediatamente es que las vocalizaciones son espectralmente muy puras y contienen, generalmente, una única frecuencia modulada.

4.2.2. Mecanismos de producción

Las ratas pueden emitir tanto vocalizaciones ultrasónicas como audibles, aunque las segundas se producen solo en un enfrentamiento directo o frente a un estímulo doloroso. Las vocalizaciones audibles presentan una estructura de armónicos similar a la de otros mamíferos y son generadas mediante la vibración de las cuerdas vocales.

Por otro lado, las vocalizaciones ultrasónicas son prácticamente tonos puros modulados lo cual sugiere un mecanismo de producción diferente. Hay relativamente pocos estudios que hayan explorado esta pregunta y seguramente los más completos sean los trabajos de Roberts y colaboradores en los años setenta (50, 51, 52, 53). En esos trabajos ellos acumulan evidencia a favor de un mecanismo puramente aerodinámico para la producción de llamadas ultrasónicas. Muestran entre otras cosas: que los sonidos son emitidos desde la laringe, que las vocalizaciones audibles y las ultrasónicas se comportan de forma diferente al reemplazar el aire por un gas más liviano (Heliox), que un silbato tipo *bird-call* con tamaños similares a los de la laringe de la rata puede generar tonos parecidos a las vocalizaciones.

Un silbato *bird-call* consiste esencialmente de dos orificios alineados de tamaños diferentes a través de los cuales fluye el aire. Inestabilidades en el flujo generan una serie de vórtices que pulsan, entran en resonancia y emiten un tono puro. No está claro, en la anatomía de la rata, qué partes del tracto aéreo servirían de orificios del silbato. Sanders et al. (54) reportan que durante la emisión de vocalizaciones ultrasónicas las cuerdas vocales se encuentran completamente aducidas dejando un pequeño orificio de 1 a 2 mm en la parte dorsal. Este podría ser uno de los orificios pero todavía queda sin respuesta la pregunta de cual podría ser el otro.

4.2.3. Clases de Vocalizaciones

Una primera categorización de vocalizaciones puede hacerse simplemente mirando el contenido espectral. Suele distinguirse entre vocalizaciones de 22 KHz y de 50 KHz (55), a veces llamadas también *vocalizaciones de baja frecuencia* y *vocalizaciones de alta frecuencia*.

Las vocalizaciones de 22 KHz son emisiones de duración larga ~ 1 s y presentan poca o ninguna modulación con una frecuencia central cercana a los 22 KHz aunque este valor puede fluctuar entre los 20 y los 25 KHz dependiendo de la cepa, del tamaño

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

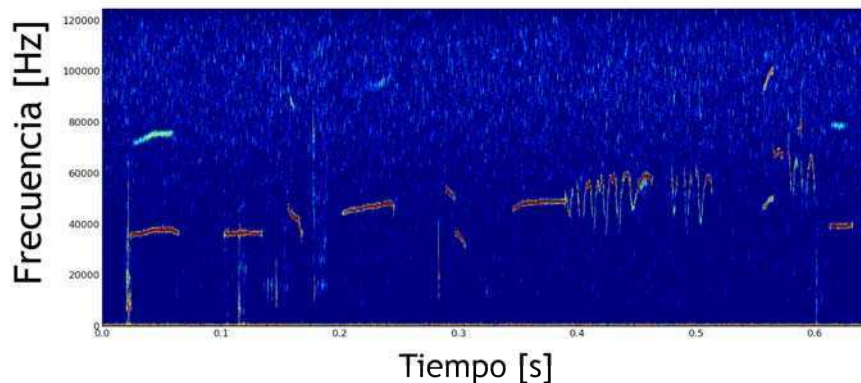


Figura 4.1: Espectros de una vocalizaciones ultrasónicas - Algunos ejemplos de vocalizaciones ultrasónicas. Si bien se trata de vocalizaciones consecutivas emitidas por una rata se borraron silencios en el medio a los fines de mostrar una mayor cantidad en un mismo gráfico.

y de la edad del animal. Una vocalización típica de la categoría de 22 KHz puede verse en la figura 4.2 panel (b).

Las llamadas vocalizaciones de 50 KHz conforman una familia más amplia y su rango de frecuencias puede ir desde los 30 KHz hasta los 80 KHz o incluso más alto en individuos jóvenes. Algunos ejemplos de esta categoría se muestran en la figura 4.2 panel (a). Las vocalizaciones de esta categoría pueden presentar una modulación amplia en frecuencia que puede ser suave o en saltos abruptos. Uno de los esquemas de clasificación de vocalizaciones más recientes y completos fue realizado por Wright et. al (56). En ese trabajo los autores elijen 14 categorías distintas de vocalizaciones de 50 KHz a partir de la inspección, a ojo, de miles de vocalizaciones. Esta categorización está resumida en una figura de su trabajo que reproducimos en 4.2. En la misma figura aparece, a modo de comparación, una vocalización de 22 KHz. Se puede ver que la riqueza de las vocalizaciones de 50 KHz es mucho mayor que las de 22 KHz.

En la siguiente sección veremos que estas dos grandes categorías cumplen roles comportamentales diferentes.

4.3. Rol Social

Las ratas son animales con un complejo comportamiento social. Viven en colonias de muchos individuos y establecen jerarquías (57). Buscan permanentemente la interacción

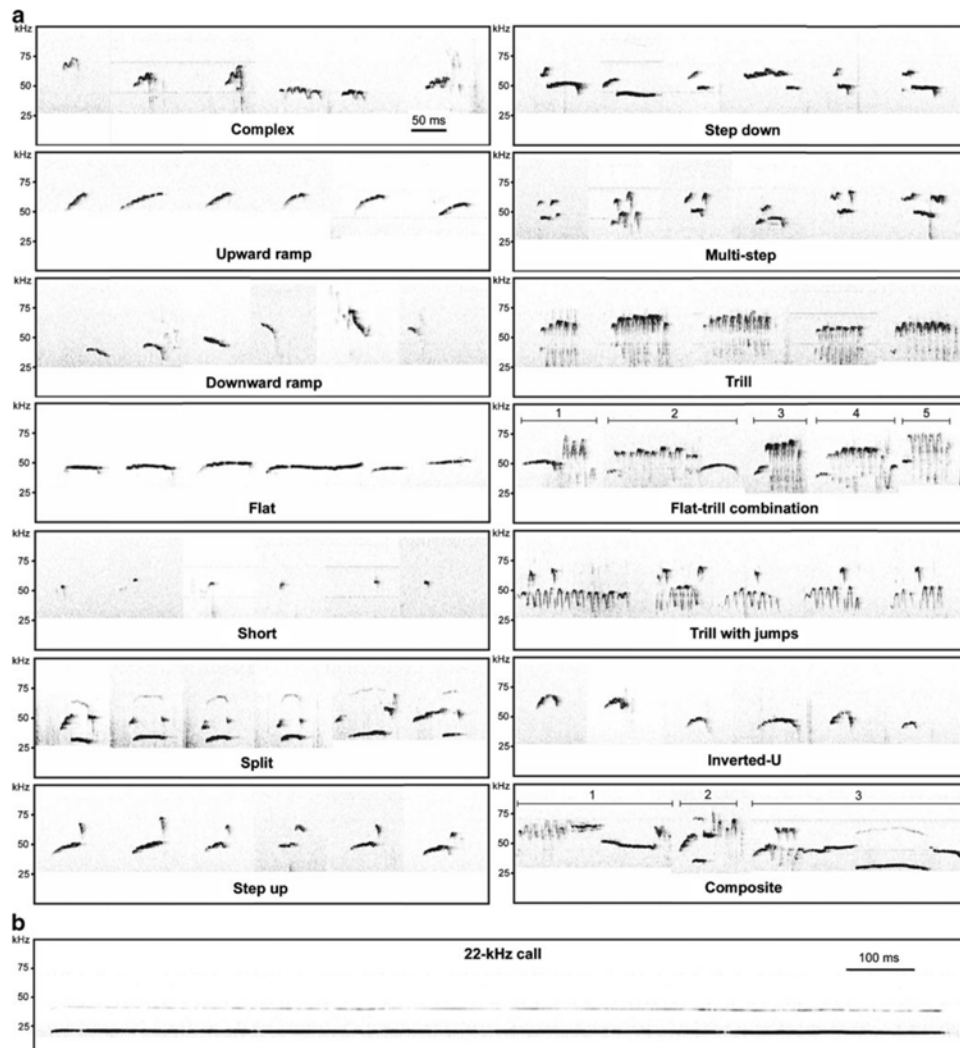


Figura 4.2: Categorías de Vocalizaciones - Categorías definidas por Wright y colaboradores (56)

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

con otros miembros de la colonia, aunque solo recientemente se comenzó a entender cuan complejos son los comportamientos empáticos y pro-sociales en la rata (58). No está claro que rol juega la comunicación vocal dentro de la organización social de la colonia pero, en una primera aproximación, las vocalizaciones parecen ser un buen indicador del estado emocional del individuo que las emite. En ese sentido, las dos grandes categorías que definimos más arriba desempeñan papeles complementarios.

4.3.1. Vocalizaciones de 22 KHz

Las vocalizaciones de 22 KHz se suelen interpretar como señales de alarma o *aflicción (distress)* debido al tipo de situaciones que evocan su producción (59). Situaciones típicas en las que una rata emite 22 KHz incluyen: la presencia distante de un predador (60), percepción del olor de un predador, la presencia de una rata dominante y agresiva (61, 62), la presencia de un humano desconocido (63), o inclusive un estímulo mínimamente desagradable (como un soplido de aire) si es presentado de manera impredecible (64). Por otro lado, la emisión de llamadas de baja frecuencia se reduce con la administración de drogas ansiolíticas (65).

Estas señales generan un sistema de alarma distribuido en la colonia ya que no es necesario que cada individuo esté atento a todas las posibles amenazas pues será informado por los demás integrantes. Para que este sistema sea efectivo las vocalizaciones de 22 KHz tienen que poder evocar respuestas defensivas (como por ejemplo *parálisis (freezing)*) en el individuo que las escucha. En efecto, Kim et al (66) muestran que las ratas, en determinadas condiciones, pueden paralizarse frente a la escucha de una señal de 22 KHz emitida por un congénere.

4.3.2. Vocalizaciones de 50 KHz

Las vocalizaciones de 50 KHz son mucho más ricas y variadas. Típicamente se evocan en contextos afectivos positivos. Ejemplos de estos contextos son: interacciones sociales no agresivas (67), durante el cortejo (68), interacciones de juego, al ser acariciados por un humano (69), al recibir drogas estimulantes (70), etc. Las ratas también emiten llamadas de 50 KHz durante la expectativa de una experiencia placentera, por ejemplo, esperando a recibir un impulso eléctrico en centros cerebrales de recompensa (71).

Los contextos sociales también producen un aumento en la emisión de vocalizaciones de alta frecuencia. Las ratas emiten un mayor número de vocalizaciones en lugares

donde previamente hubo otras ratas. Más aún, el número de vocalizaciones que emiten es directamente proporcional a la cantidad de individuos distintos que hayan estado en ese mismo espacio antes (67).

Si bien vimos que dentro de las vocalizaciones de alta frecuencia se pueden definir hasta 14 categorías distintas, no es evidente que dichas categorías tengan un correlato comportamental. Generalmente, no se consideran las 14 categorías sino que se divide a las vocalizaciones en chatas (*flat*) o de Frecuencia Modulada (FM). Burgdorf et al. (72) muestran que las ratas se autoadministran voluntariamente grabaciones de vocalizaciones de 50 KHz de frecuencia modulada pero no chatas. Esto los hace concluir que no todas las llamadas de alta frecuencia se corresponden a un contexto afectivo positivo sino que eso solo es válido para las de FM. Sin embargo, Wohn et al. (73) muestran que la presentación tanto de vocalizaciones FM como de tonos puros provoca que las ratas se acerquen al parlante que las está emitiendo, aparentemente contradiciendo lo anterior. Hay experimentos que muestran que es posible que las ratas emitan vocalizaciones de 50 KHz en contextos neutros o, inclusive, levemente negativos. Por ejemplo, al ser separados de sus compañeros de jaula, tanto la rata que fue retirada como el individuo que queda en ella emiten vocalizaciones de alta frecuencia, mayormente sin modulación de frecuencia (74).

4.4. Circuitos Neuronales

Las ratas comparten con otros mamíferos parte de los circuitos neuronales responsables del comportamiento vocal motor (75). Las motoneuronas que enervan los músculos de la laringe se encuentran en el *Núcleo Ambiguo* (NA) dentro del bulbo raquídeo. Este núcleo tiene conexiones recíprocas con la *Formación reticular lateral* (LRF) donde se presume que está el generador central de patrones. La evidencia para esto es que registros de una sola célula en esa región encuentran que la actividad neuronal sigue a la estructura temporal fina de cada vocalización (76) a pesar de que no enerva directamente a la laringe. LRF tiene también conexiones recíprocas con la *Sustancia Gris Periacueductal* (PAG) una estructura que media la iniciación de una vocalización. Lesiones en esta región producen mutismo en ratas, gatos, perros e inclusive en pacientes humanos(77). Tanto la estimulación eléctrica como química de PAG evoca vocalizaciones muy similares a las naturales (78) mientras que la estimulación en LRF produce

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

vocalizaciones deformadas. Más aún, la actividad neural en PAG no sigue los detalles finos de las vocalizaciones sino que la mayoría de las células presenta actividad fásica sincronizada al inicio de la vocalización o tónica durante la duración de la misma (79).

Toda esta evidencia parece indicar que LRF cumple el rol de un generador central de patrones mientras que PAG desempeña un papel en la iniciación de las vocalizaciones. PAG a su vez, recibe conexiones directas de distintas estructuras del sistema límbico relacionadas con estados motivacionales diferentes(80).

Dado que, como vimos, las vocalizaciones de 22 KHz y las de 50 KHz están asociadas a contextos afectivos diferentes es esperable que se pueda afectar diferenciadamente la tasa de emisión de los dos tipos de vocalizaciones manipulando el funcionamiento de ciertas estructuras del sistema límbico. En efecto, la estimulación del sistema colinérgico con origen en el *Núcleo Tegmental Laterodorsal* (LTN) induce en la rata un estado defensivo y evoca intensamente llamados de 22 KHz (81). Por otro lado, la activación del sistema dopaminérgico de la rata, con origen en el *Área Tegmental Ventral* (VTA) aumenta la actividad locomotora, el comportamiento de exploración e induce vocalizaciones de 50 KHz (70, 82, 83).

4.5. Control Voluntario

Vimos que PAG juega un rol en la iniciación de vocalizaciones, pero no está claro el grado de control voluntario que hay en la iniciación de esas vocalizaciones. PAG recibe conexiones de la Corteza Cingulada Anterior (ACC) una región que presumiblemente juega un papel en la iniciación voluntaria de vocalizaciones. La evidencia para esto proviene de estudios en monos (84) donde los entrenan para responder con vocalizaciones específicas a ciertos estímulos y luego ven que esta habilidad se pierde al lesionar ACC. Si bien está demostrado que hay conexiones desde ACC hacia PAG en la rata (85) la evidencia comportamental para el uso voluntario de vocalizaciones es poca y contradictoria. Hay cuatro trabajos publicados, de los que tengamos conocimiento, en los que hay reportes ya sea positivos o negativos de condicionamiento de vocalizaciones. Todos los trabajos son acerca de vocalizaciones audibles. Los primeros resultados positivos corresponden a los trabajos de 1943 de Cowles Pennington (86) y 1946 de Herbert (87). En esos dos trabajos se describen protocolos exitosos de condicionamiento **clásico** donde usan un shock eléctrico como estímulo incondicionado y un tono como

estímulo condicionado. Posteriormente el trabajo de Mowrer y colaboradores de 1948 (88) muestra resultados negativos frente a un protocolo de condicionamiento **operante**. Finalmente, el último de los trabajos, el de Lal de 1967 (89) muestra resultados positivos en algunas de las ratas entrenadas bajo un protocolo operante. No hay ningún trabajo publicado sobre el condicionamiento de vocalizaciones ultrasónicas.

4.6. Respiración

Un aspecto relevante y a tener en cuenta en el estudio de vocalizaciones es el ritmo de respiración que presenta la rata. Éste será distinto según la rata se encuentre en un estado de reposo o durante una exploración activa de la escena olfativa (*sniffing*). La respiración en reposo tiene un ritmo cercano a los 2 Hz mientras que durante el *sniffing* la respiración se acelera al rango de los 8 a 12 Hz (90). Dado que se emite una vocalización por exhalación, las mismas suelen venir agrupadas en ráfagas (*bouts*) a la frecuencia de **sniffing**. Estos elementos resultarán importantes a la hora de segmentar vocalizaciones a partir de audio continuo.

4.7. Dispositivo Experimental

Dadas las características etológicas que describimos en las secciones anteriores decidimos armar una arena que nos permita grabar vocalizaciones de un único individuo pero en compañía de un congénere. Para entender sistemas de comunicación social, es importante contar con registros en los que es posible asignar las vocalizaciones al individuo que las emitió. Esto es un desafío particular, en el caso de la rata, por su tamaño y por la intensa interacción física de contacto que despliegan. Es por eso que decidimos construir una arena para dos ratas en la que estén separadas pero puedan interactuar vocalmente. De esta manera, tenemos un dispositivo experimental que provee un contexto favorable para la emisión de vocalizaciones de 50 KHz y que conserva ciertas características etológicas de la interacción pero, a la vez, permite asignar las vocalizaciones a cada individuo. En las siguientes secciones describimos los elementos necesarios para la construcción de dicha arena y el protocolo utilizado en los experimentos.

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

4.7.1. Hardware

La arena de grabación está compuesta por una jaula de alambre recubierto (150 cm de largo, 50 cm de ancho, 70 cm de alto) dividida en dos partes con una separación de 10 cm entre ellas. Un esquema puede verse en la figura 4.3. Pueden verse en esa misma figura los demás componentes. Cada mitad de la arena posee un micrófono ultrasónico (Avisoft-Bioacoustics CM16/CMPA) en la parte de arriba y una cámara USB infrarroja (Microsoft LifeCam 3000 con el filtro de infrarrojo quitado). Ambas mitades se encuentran semi-separadas (la parte más baja se encuentra descubierta) por espuma acústica de manera tal de minimizar el ruido cruzado entre ambos micrófonos pero permitiendo que las vocalizaciones pasen entre una y otra mitad de la arena.

Los registros se toman a oscuras y por lo tanto hace falta iluminar la arena con luz infrarroja para poder grabar video. Para esto, se coloca un array de leds infrarrojos por encima de todo el dispositivo. Para evitar reflexiones sonoras con los objetos de alrededor, se coloca espuma acústica rodeando la jaula a unos 20 centímetros de distancia (para evitar que las ratas la muerdan). A su vez, con el fin de minimizar el ruido introducido por la caminata de los animales decidimos usar un tipo específico de aserrín para el piso (Aspen Chips).

La digitalización de señal de los micrófonos se realiza mediante una placa de adquisición (National Instruments PCIe-6320). Para poder sincronizar el video con el audio se envía, usando la misma placa NI, una señal de sincronía a dos leds infrarrojos que aparecen en el campo de visión de cada cámara. Todo el dispositivo se encuentra adentro de una habitación insonorizada. Las PCs de registro se ubican afuera de la habitación dado que *coolers* y otros dispositivos internos de las computadoras producen ruido en el rango ultrasónico.

4.7.2. Software

Describiremos el software para el análisis del audio con detalle en el próximo capítulo pero en esta sección haremos un breve racconto del procesamiento de la señal de video y su señal de sincronía.

Usando la librería de acceso libre Open CV (91) implementamos una pequeña rutina que permite extraer la posición y orientación de la rata en cada cuadro del video. Dado que el aserrín ofrece un buen contraste contra el color oscuro de la rata es posible

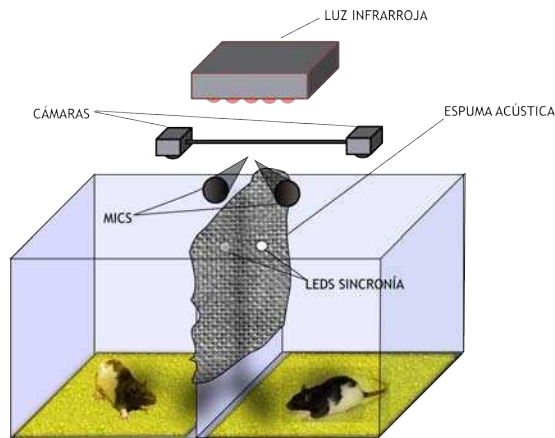


Figura 4.3: Arena de Grabación - Esquema con los diferentes componentes de la arena de grabación

identificarla usando simplemente un umbral de brillo. En la figura 4.4 mostramos un cuadro original del video y a su lado un cuadro con el umbral aplicado. Allí los pixels más claros que cierto valor se convierten en blanco y los pixels más oscuros que ese mismo valor se convierten en negro. Luego, es posible buscar la componente conexa negra más grande que haya en la imagen y ajustarla por una elipse. Ésta nos dirá la posición y orientación de la rata.

Por otro lado necesitamos mantener la sincronía entre el audio y el video. Para eso enviamos una señal generada por la placa de adquisición (y que por lo tanto podemos garantizar que estará en sincronía con el audio grabado) hacia dos leds que están en el campo de visión de cada cámara. En el cuadro de video de la figura 4.4 se puede ver destacado el led de sincronía. Debajo, en el panel C, mostramos la señal de sincronía usada. La señal consiste en una sucesión de pulsos de dos largos distintos. Los pulsos de 500 ms actúan de separadores y se encuentran marcados en la figura. Los pulsos cortos (250 ms) indican el número de separador del que se trata en binario. De esta forma es posible sincronizar audio y video aunque se haya perdido alguna parte del registro.

4.7.3. Protocolo

Para los registros se usaron ratas de la cepa Long Evans, adultos jóvenes (circa 3 meses). El experimento que se analiza en el capítulo siguiente fue realizado en 6 ratas macho (V1...V6) que vivían separadas en dos jaulas de 3 y 3 (V1-V3 / V4-V6). Las sesiones de grabación fueron de 10 minutos cada una y estaban divididas en sesiones de

4. VOCALIZACIONES ULTRASÓNICAS EN RATAS.

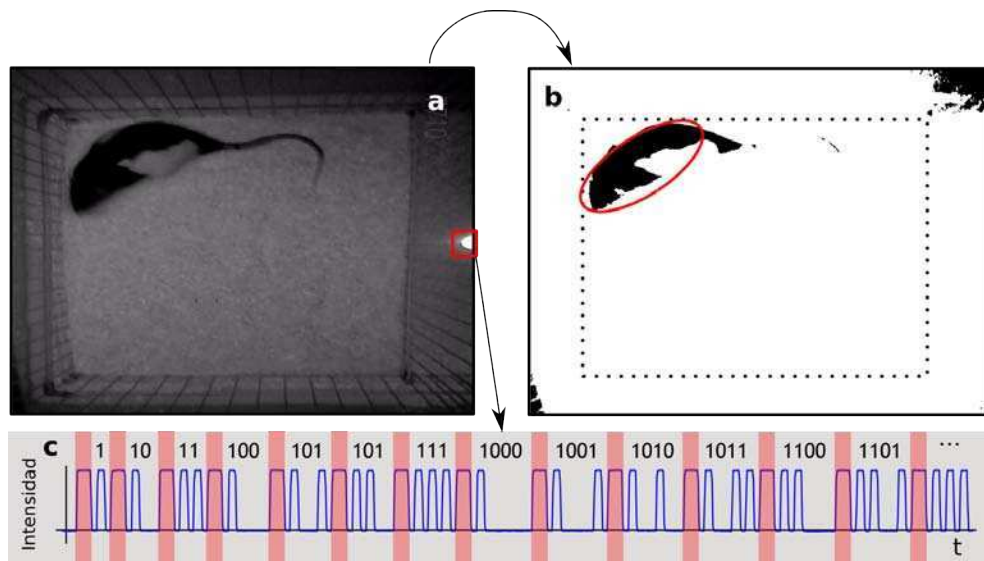


Figura 4.4: Análisis de los videos - En el panel (a) se muestra, a modo de ejemplo, un cuadro de video. En el panel (b) se muestra el mismo cuadro pero con un umbral de brillo donde la rata se destaca claramente como un objeto oscuro. Se muestra también el ajuste por una elipse. En el panel (c) se presenta la señal de sincronía enviada al led infrarrojo.

interacción y sesiones solitarias ordenadas en forma aleatoria. Se registraron 15 sesiones de pares (una para cada par posible) y 6 sesiones solitarias para cada rata. Las sesiones fueron registradas a lo largo de 3 días. Entre sesión y sesión se reemplaza todo el aserrín de la arena y se limpia con una solución de agua y alcohol al 70 %. Esto se hace para minimizar la presencia de olores de las ratas que estuvieron anteriormente. Bajo un esquema similar contamos también con registros de pares Macho-Hembra que son discutidos cuando analizamos interacciones vocales entre individuos.

5

Clasificación automática y estructura de vocalizaciones

El esquema de clasificación automática que presentamos en este capítulo parte de la clasificación de Wright et al. que estuvimos discutiendo en el capítulo anterior. Nuestro objetivo al desarrollar esta herramienta es doble. Por un lado, nos proponemos estudiar en que medida las clasificaciones de Wright emergen de las distribuciones de las propiedades acústicas de las llamadas. Por otro, la inspección visual de cada una de las vocalizaciones es engorrosa y proclive a sesgos. Desarrollar algoritmos que permitan agilizar y sistematizar esta tarea permitiría el estudio a gran escala de corpus de vocalizaciones ultrasónicas. En este capítulo nos abocamos precisamente a esa empresa. Completa la herramienta, la utilizamos para categorizar automáticamente las vocalizaciones de la arena de diálogos y estudiamos cómo las diferentes categorías se estructuran en el tiempo. Hacia el final del capítulo analizamos, también, la existencia de interacción o no entre las vocalizaciones de diferentes individuos.

5.1. Estrategia general

Dado que, como vimos, las vocalizaciones presentan, en su gran mayoría, únicamente **una** frecuencia para cada tiempo - es decir, son básicamente tonos puros modulados- podemos reducir drásticamente la dimensionalidad del espacio a estudiar si en lugar de conservar la información espectral completa, nos quedamos solamente con la frecuencia pico para cada tiempo. Esta idea está representada esquemáticamente en la figura 5.1.

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

Allí también se representan las etapas de procesado posteriores. Una vez obtenida la curva definimos propiedades de la misma que sirvan para separar sus tramos en el tiempo. En la siguiente sección daremos una definición precisa de estas propiedades. Finalmente, en el último cuadro de la misma figura mostramos como una vocalización termina siendo representada por una curva en el espacio de propiedades. Regiones en el espacio de propiedades corresponderán a clases de segmentos de vocalización.

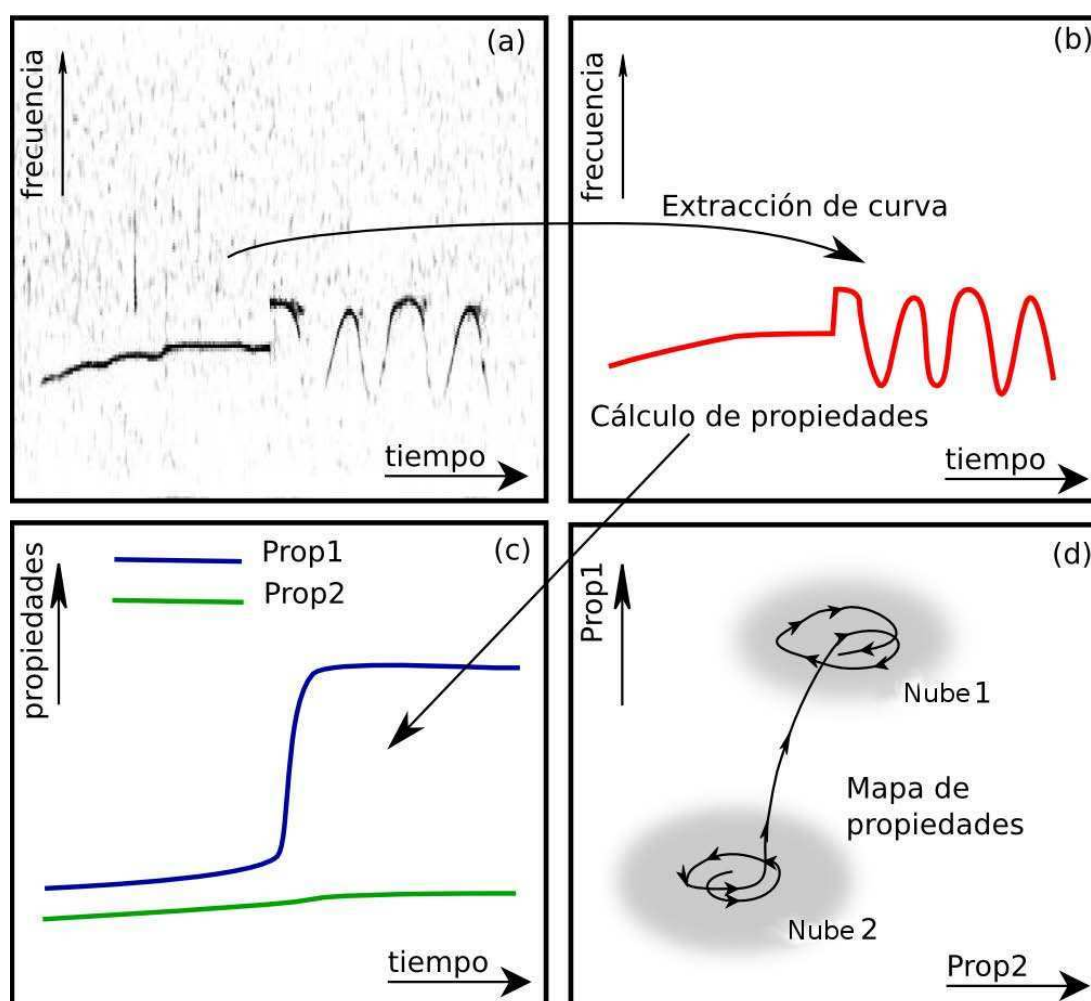


Figura 5.1: Estrategia general para la categorización de curvas. - (a) Cálculo del espectrograma de reasignación (92) de una vocalización. (b) Extracción de curva a partir del espectrograma. (c) Cálculo de propiedades de la curva en función del tiempo. (d) Elaboración del mapa de propiedades para categorizar segmentos de curva.

Una vez que tengamos clasificados los segmentos, estudiaremos las transiciones entre clases que ocurren **dentro** de una vocalización. Esto nos permitirá identificar ciertas

secuencias estereotipadas y compararlas con las categorías definidas ad-hoc por Wright et al. Finalmente analizaremos la estructura en una escala temporal más larga mirando transiciones de categorías **entre** vocalizaciones.

5.2. Detección y Extracción de propiedades

El dato crudo a partir del cual parte todo este análisis es el clip de audio grabado a 250 kHz en la arena de diálogos que describimos en el capítulo anterior. En una primera instancia tenemos que poder identificar los momentos en los que cada rata vocaliza. Lo más sencillo sería computar el poder total de la señal tiempo a tiempo y poner un umbral, un volumen mínimo a partir del cual consideramos que la rata está vocalizando. Sin embargo, esa estrategia no es sensible al contenido espectral y por lo tanto terminará identificando como vocalización a cualquier ruido que cruce el umbral (por ejemplo el ruido de la rata caminando o saltando sobre el aserrín, mordisqueando los barrotes, etc).

Una característica saliente de las vocalizaciones, vimos, es su monotonalidad. Es decir que tiempo a tiempo se aproximan bien por una única componente armónica. Esto resulta suficiente para distinguirla de los demás ruidos presentes en la jaula. Usaremos entonces la entropía de la distribución de poder como principal propiedad para separar señal de ruido. Definamos esto un poco más rigurosamente.

Si llamamos P_{ij} a la potencia de la señal en la frecuencia f_i al tiempo t_j - es decir, que la matriz P es el espectrograma de la onda sonora- definimos entonces la entropía para cada tiempo t_j :

$$H_j = - \sum_i^N p_{ij} \log(p_{ij}) \quad (5.1)$$

donde $p_{ij} = \frac{P_{ij}}{\sum_i P_{ij}}$. Recordemos que esta expresión valdrá 0 si toda la potencia está concentrada en una única frecuencia f_i^* y valdrá $\log(N)$ si la potencia está uniformemente distribuida en todas las frecuencias. Esto nos permitirá establecer un umbral por debajo del cual consideraremos que se produjo una vocalización. Nos resta explicitar como calcularemos P_{ij} .

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

5.2.1. Cálculo del espectro.

La forma usual en la que se calcula un espectrograma es haciendo uso de la transformada de Fourier en una ventana temporal móvil. De esta manera, se puede construir un gráfico de intensidad en función de la frecuencia para cada tiempo. En este esquema solo se usa el valor de la amplitud y se descarta la información de la fase. Desde los años 70 existen propuestas para aumentar la precisión del espectrograma reasignando cada punto del espacio (t, ω) (tiempo, frecuencia) a la frecuencia y tiempo instantáneos (92, 93, 94). A esta familia de técnicas se la conoce como espectrogramas de reasignación. Este esquema si bien no aumenta la resolución (no podremos distinguir dos señales que estén mas cerca que el límite de incerteza) permite localizar señales aisladas con mucha mayor precisión. Estas técnicas son particularmente buenas para observar modulaciones temporales de frecuencia. En la figura 5.2 presentamos la misma vocalización con iguales tamaños de ventana y solapamiento para un espectrograma común y para un espectrograma de reasignación. Los espectros de reasignación fueron calculados usando la librería de software libre LibTfr (95). A partir de ahora siempre que nos refiramos al espectro de la señal será calculado usando el espectrograma de reasignación.

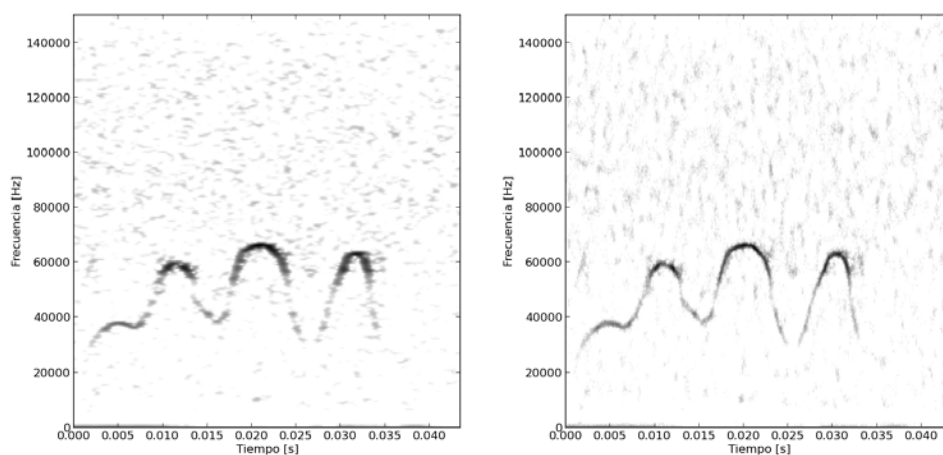


Figura 5.2: Ejemplos de espectrogramas - IZQ: Espectrograma convencional ; DER: Espectrograma de reasignación.

5.2 Detección y Extracción de propiedades

Parámetro	Valor
H_{th1}	4.5
H_{th2}	4.8
WIN	512
$STEP$	30

Cuadro 5.1: Parámetros de los espectrogramas y umbral para la detección de vocalizaciones en el archivo de audio. H_{th1} es el primer umbral de entropía; H_{th2} el segundo; WIN es el tamaño de la ventana del espectrograma y $STEP$ el número de paso.

5.2.2. Segmentación de vocalizaciones

Ahora sí, estamos en condiciones de segmentar el archivo de audio continuo en vocalizaciones. Para esto, como dijimos en la sección anterior, calculamos la entropía usando el espectrograma que definimos recién. Luego definimos un umbral de entropía por debajo del cual consideramos que hubo una vocalización. En la figura 5.3, en panel del medio puede verse la señal de entropía para una vocalización junto con dos umbrales. Usamos un sistema de doble umbral en el cual un umbral más estricto (entropía más baja) se requiere para decidir si hubo o no una vocalización pero luego un umbral más laxo define cuales serán los bordes de la vocalización. Cuales sean los valores de los umbrales dependerá de los parámetros con que se haya calculado el espectrograma. En la tabla 5.1 resumimos los valores usados para el cálculo de espectro y umbrales de entropía.

Si definimos que los momentos en los que la entropía no supera el umbral son silencios y que los silencios separan vocalizaciones resultaría que el segmento de audio que mostramos en la figura 5.3 quedaría separado en cuatro vocalizaciones distintas. Sin embargo, vamos a definir que si dos segmentos del audio donde decidimos que ocurrió una vocalización están más cerca que 20 ms diremos que son parte de la misma vocalización. Una de las razones para esto es que de esa forma nuestra segmentación automática coincide con la manera usual de segmentar vocalizaciones a ojo según se reporta en la literatura. Pero por otro lado, hay razones fisiológicas para usar este criterio. Si observamos un histograma de la duración de los silencios, según los definimos al principio de este párrafo, veremos algo llamativo (ver figura 5.4). Allí vemos que las duraciones de los silencios presentan mucha estructura. Hay un pico muy marcado cerca de la duración cero que sale de la escala del gráfico y luego dos picos menores; uno se encuentra cerca de los 100 ms y el otro cerca de los 200 ms. Esta estructura

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

es la que impone, esencialmente, la respiración. Según vimos en el capítulo anterior, las ratas poseen un modo de respiración estereotipado durante fases de exploración que llamamos *sniffing*. Ésta es una respiración rápida, de entre 8 y 12 Hz, que se corresponde con tiempos entre respiraciones de aproximadamente 100 ms. Entonces, los picos que se observan en el histograma, de 100 ms y 200 ms, corresponden a silencios de una respiración y dos respiraciones respectivamente. Por otro lado, el pico cercano a cero contiene silencios que están dentro de una misma exhalación. Vemos, entonces, que hay un margen para definir un tiempo de corte (entre 20 ms y 60 ms) que nos permite agrupar todos los sonidos que hayan sido producidos en una exhalación como una única vocalización.

Finalmente, para definir la curva que representará a la vocalización, basta tomar el valor de la frecuencia pico en el espectrograma para cada tiempo solo para los puntos en que la entropía esté por debajo del umbral más estricto. A los puntos que cruzaron el umbral más estricto, se les permite relajar y contagiarse a sus vecinos hasta llegar al umbral más laxo. En último panel de la figura 5.3 mostramos con línea gruesa cuales serían los puntos seleccionados por este método.

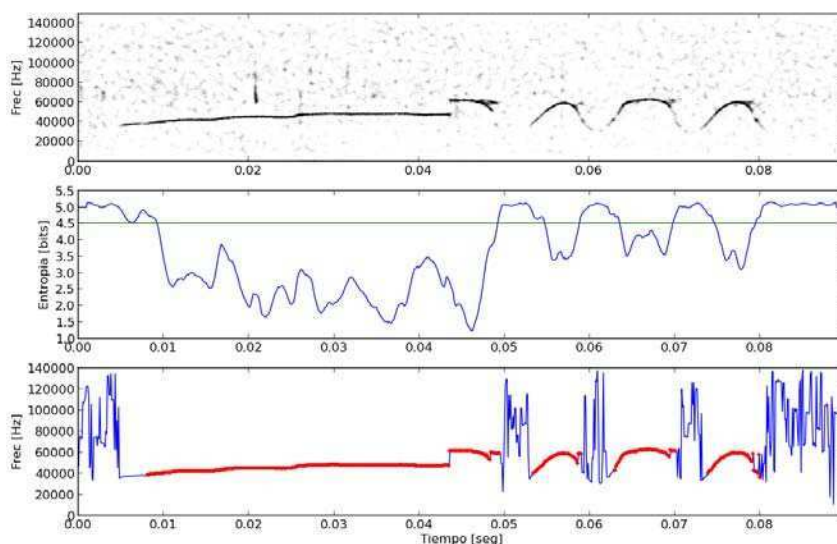


Figura 5.3: Ejemplo de extracción de una curva -

Ahora que ya llevamos las vocalizaciones a una representación más sencilla, nos resta

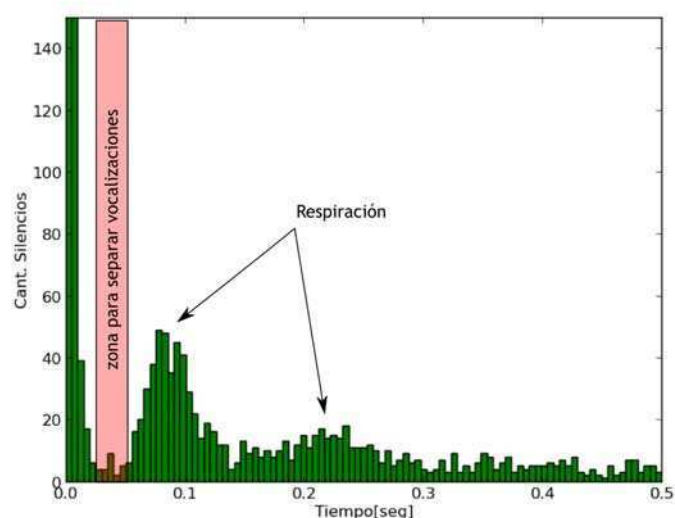


Figura 5.4: Histograma de duración de silencios - Los silencios son definidos como momentos donde la entropía no llega a cruzar el umbral. Se observan dos picos que corresponden a haber dejado una y dos respiraciones entre vocalizaciones, respectivamente.

definir propiedades para seguir avanzando con el esquema general que planteamos al principio. Es decir, nos encontramos a punto de definir el panel (c) de la figura 5.1.

5.2.3. Propiedades de la curva

Utilizaremos dos propiedades para caracterizar a las curvas: la frecuencia instantánea y la modulación. La frecuencia instantánea no es otra cosa que la curva misma. Es decir, qué frecuencia está emitiendo la rata momento a momento. La modulación, por otro lado, corresponderá a la amplitud de la variación de frecuencias. Dado que no siempre se registra el ciclo completo en una vocalización modulada, pues la intensidad disminuye mucho para las frecuencias bajas (ver por ejemplo la figura 5.3), no es posible simplemente restar el máximo y el mínimo de la señal para obtener la amplitud. Para remediar esto, aprovecharemos el hecho de que las variaciones en la frecuencia tienen una frecuencia característica estable de alrededor de 100 Hz. Mediante un ajuste lineal en una ventana móvil, podemos reconstruir la señal original y, a partir del ajuste, obtener los valores de la modulación.

Sin embargo, antes de poder realizar el ajuste, hace falta hacer una pequeña corrección en las señales. Como describimos en el capítulo anterior, las vocalizaciones

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

ultrasónicas suelen presentar saltos abruptos de frecuencia. Para ajustar solo partes continuas de la señal, detectaremos saltos de frecuencia mayores que un cierto valor y los usaremos como bordes. Es decir, solo ajustaremos curvas entre saltos o, dicho de otra forma, la ventana móvil del ajuste no es capaz de cruzar saltos.

La función que ajustaremos será la siguiente:

$$f(t) = A + Bt + C\cos(2\pi\nu t) + D\sin(2\pi\nu t) \quad (5.2)$$

A partir de esa fórmula, definiremos la modulación como $R = \sqrt{C^2 + D^2}$. El parámetro ν no se ajusta sino que se fija de ante mano. En principio ν podría variar para cada individuo pero nosotros encontramos que un valor de $\nu = 95\text{Hz}$ arrojó buenos ajustes para todos los animales analizados. La elección de este valor de ν fue hecha observando la distribución espectral de potencia para las curvas de frecuencia en función del tiempo para cada uno de los individuos (ver figura 5.5).

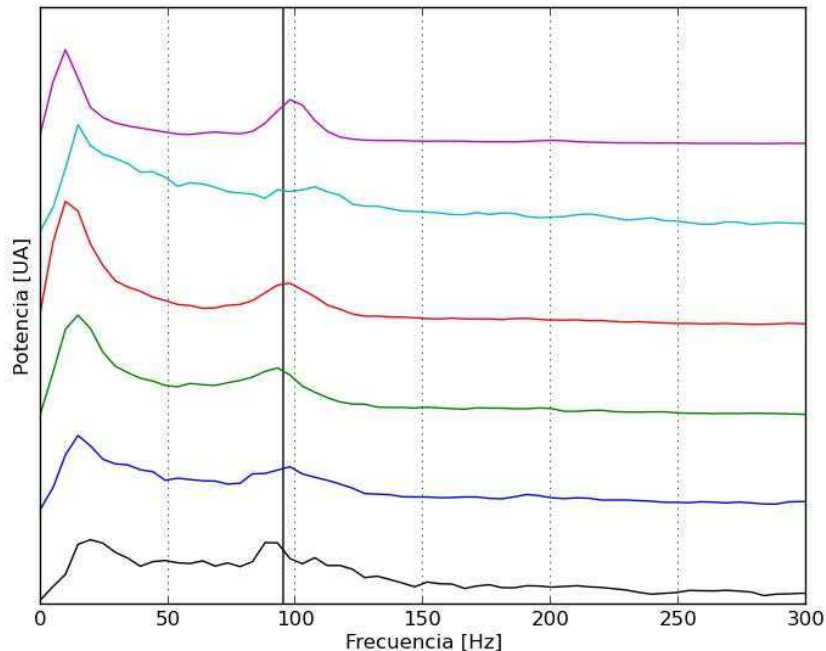


Figura 5.5: Distribución espectral de potencia de las curvas - Las distribuciones de potencia están re-escaladas y sumadas a una constante de manera tal de separarlas en el gráfico y poder compararlas. Cada curva corresponde a un animal distinto. La línea vertical marca los 95 Hz

5.2 Detección y Extracción de propiedades

En la figura 5.6 mostramos un ejemplo de curva ajustada junto con los valores de amplitud de modulación obtenidos.

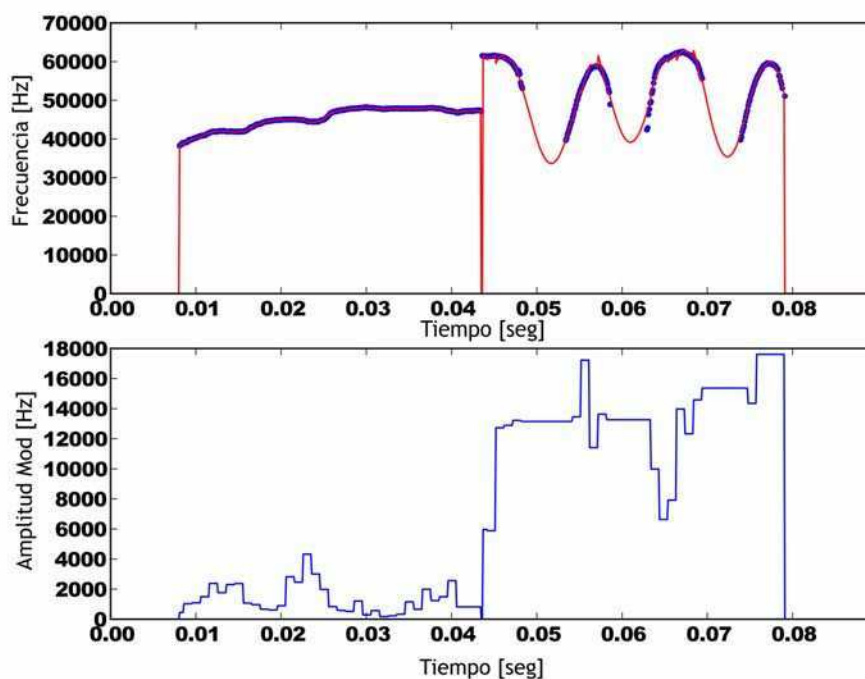


Figura 5.6: Ejemplo de curva con su regresión - SUPERIOR- Curva extraída junto con su ajuste por regresión de la ecuación 5.2 en rojo. Se puede ver como se reconstruye la oscilación en las partes en las que no hay sonido. INFERIOR- Valor de la amplitud de modulación en función del tiempo.

5.2.4. Mapa de Propiedades

Una vez que tenemos definidas y medidas las propiedades de las curvas podemos poner todas las vocalizaciones juntas en un mapa como en el panel (d) de la figura 5.1. El resultado puede verse en la figura 5.7. Allí presentamos los mapas para seis animales distintos (V1 hasta V6). Hay tres regiones que resultan distinguibles: una región de modulación elevada y predominantemente altas frecuencias (T) ; una región de baja modulación y alta frecuencia (F) y una región de baja modulación y baja frecuencia (f). No todas las regiones están presentes con igual intensidad en todos los individuos. Algunos producen predominantemente vocalizaciones del tipo T, otros del tipo F y otros producen ambas. Las vocalizaciones del tipo f son las menos observadas.

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

Nótese también, que V1 no presenta una separación clara entre sus nubes T y F. Ahora bien, ¿Qué representan estas nubes? ¿Corresponden de alguna manera a las categorías definidas por Wright y colaboradores?

Para responder estas preguntas presentamos algunos ejemplos representativos de cada una de las regiones para la rata V6 (ver figura 5.8). En los distintos paneles A-D mostramos los siguientes ejemplos: A) una vocalización sin modulación y de alta frecuencia (*flat* o F); B) una vocalización de frecuencia modulada (*Trill* o T); C) una transición de F a T; D) una vocalización que tiene una parte en la nube f. Típicamente (y como se observa en este caso), las transiciones desde y hacia *f* se realizan mediante saltos en frecuencia. El hecho de que existan nubes separables de vocalizaciones, en términos de ciertas propiedades temporo-espectrales, nos sirve para ganar confianza en la categorización ad-hoc hecha por Wright y colaboradores.

Si volvemos a revisar la figura 4.2 con las categorías de Wright et al, vemos que las regiones de los mapas de propiedades, junto con sus posibles combinaciones, nos permiten detectar varias de esas categorías: *flat*, *trill*, *split*, *flat-trill*, *composite*. Vemos también que la categoría *complex* debería aparecer como una nube de modulación intermedia, pero en nuestros mapas no se observa. Esto puede deberse a distintas razones: 1) Podría ser que no formen una categoría discreta en sus propiedades sonoras 2) Podría ser que sí formen categorías discretas, pero nuestros análisis no son capaces de separar las nubes 3) Podría ser que su producción es relativamente baja y no aparecen como una nube clara. Hay que tener en cuenta que según Wright y colaboradores, *flat*, *trill*, y *flat-trill* representan el 70% de todas las vocalizaciones.

Otra pregunta razonable, observando la misma figura, es si podemos distinguir *downward-ramps*, *upward-ramps* y *flats* que, para nosotros, formarían todos parte de la nube F. Para esto, podemos aprovechar el parámetro *B* (pendiente) de la ecuación 5.2 que usamos para el ajuste. Presentamos un histograma del parámetro *B* en la figura 5.9. Lo primero que se observa es que hay muchos más eventos con pendiente positiva que negativa. También se observa cierta estructura (destacada con flechas) aunque no es posible identificar tres picos que correspondan a *downward-ramp*, *flat* y *upward-ramp*.

Para ganar confianza en la identificación de las nubes con elementos distintos de la vocalización podemos hacer histogramas de duración para la nube F y para la nube T. Esto es lo que mostramos en la figura 5.10. Ambos poseen estructuras marcadamente distintas. La nube T posee ciertas duraciones estereotipadas y que son consistentes con

5.2 Detección y Extracción de propiedades

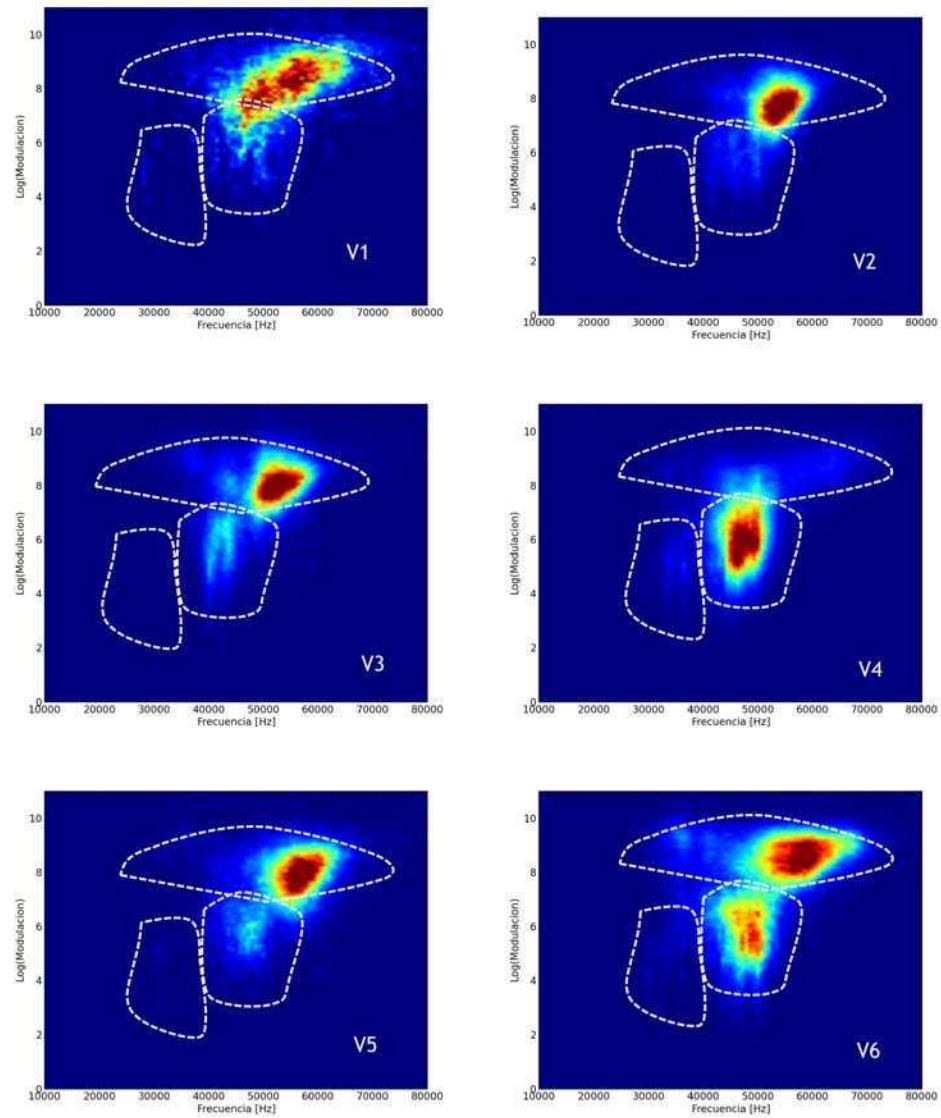


Figura 5.7: Mapas de propiedades - Mapas de propiedades para seis animales distintos. Se encuentran destacadas 3 regiones distinguibles que corresponden a formas distintas de vocalización.

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

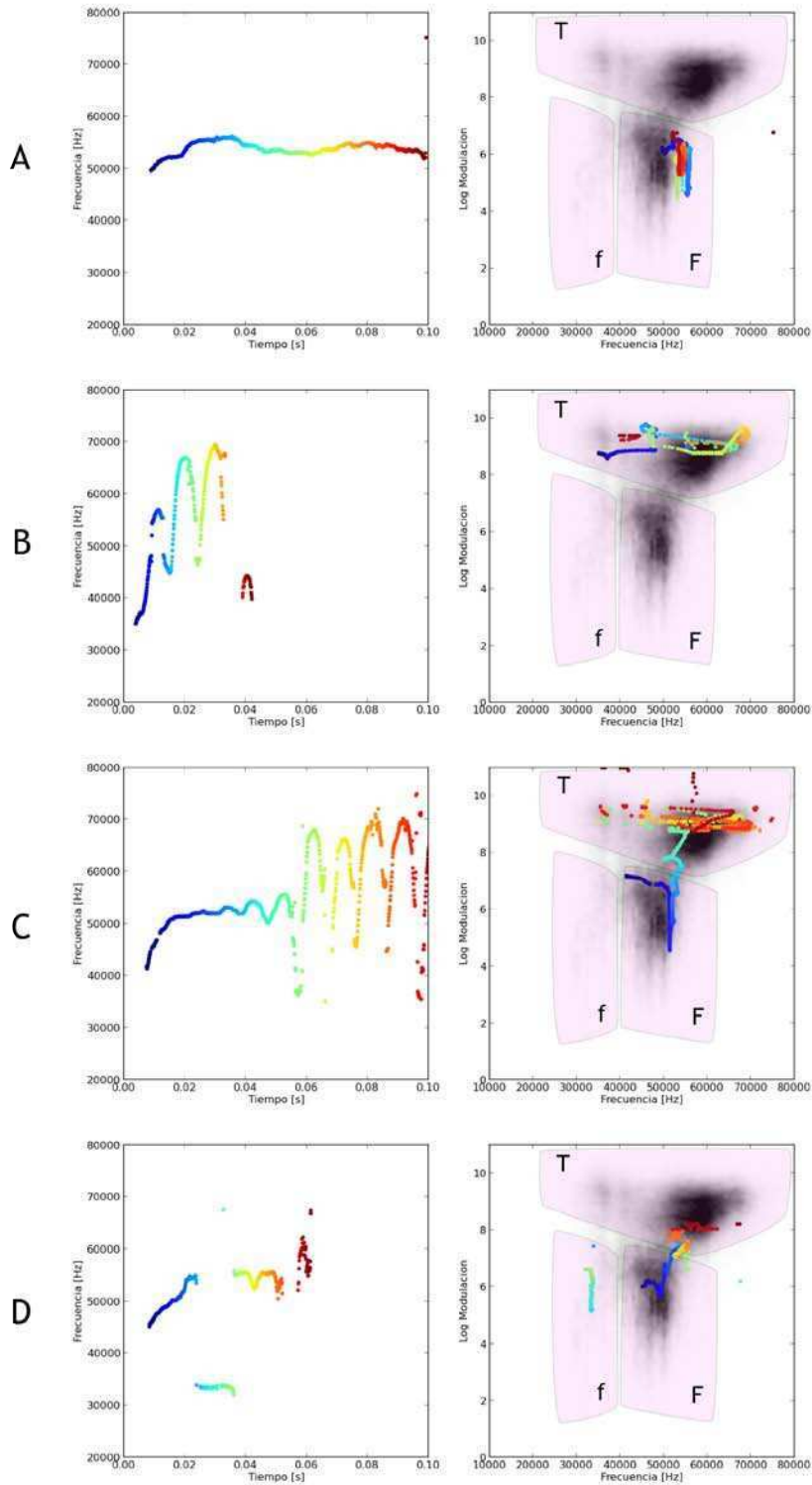


Figura 5.8: Excursiones de vocalizaciones en el espacio de propiedades - Izquierda: Curvas de frecuencia vs tiempo para distintos tipos de vocalizaciones. Derecha: Las mismas vocalizaciones pero en el espacio de propiedades. El color indica el tiempo para hacer más fácil la comparación entre ambos espacios.

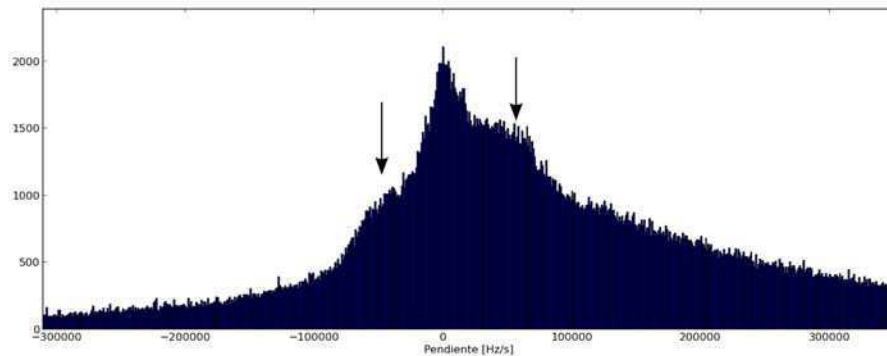


Figura 5.9: Histograma del parámetro B en el ajuste - Lo primero que se observa es que hay muchos más eventos con pendiente positiva que negativa. También se observa cierta estructura (destacada con flechas) aunque no es posible identificar tres picos que correspondan a *downward-ramp*, *flat* y *upward-ramp*

incluir 1, 2, 3, ... ciclos de una oscilación. Aparece también un pico correspondiente a media oscilación. En cambio, la nube F posee un pico claro para duraciones cortas y luego una distribución de campana centrada cercana al cero. El hecho de que al separar por propiedades espectrales también hace que se separen en su estructura de duraciones nos permite ganar confianza en nuestro esquema de clasificación. También, aparece un hecho llamativo y es que los ciclos de T aparecen de manera discreta.

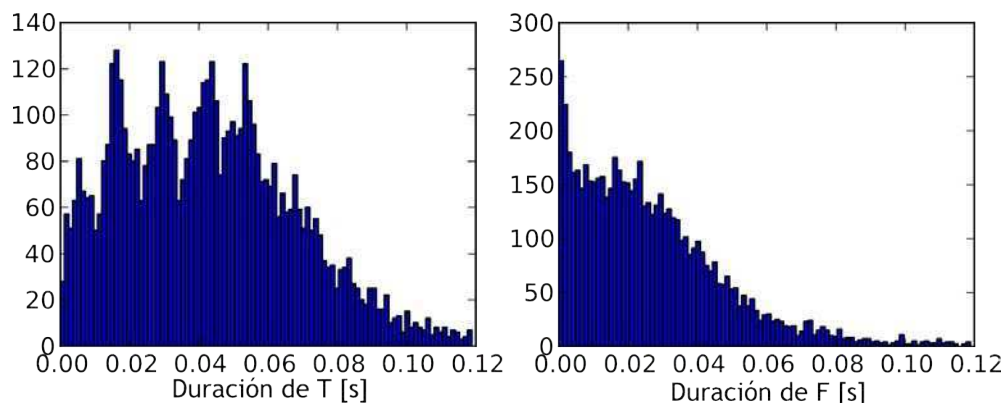


Figura 5.10: Histogramas de duración para las nubes F y T - Ambos poseen estructuras marcadamente distintas. La nube T posee ciertas duraciones estereotipadas y que son consistentes con incluir 1, 2, 3, ... ciclos de una oscilación. Aparece también un pico correspondiente a media oscilación. En cambio, la nube F posee un pico claro para duraciones cortas y luego una distribución de campana centrada cercana al cero.

En nuestro esquema, entonces, una vocalización queda reducida a un código de

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

letras que identifica por cuál región del espacio de propiedades pasa, y en qué orden. Por ejemplo, para las vocalizaciones de la figura 5.8 serían los códigos: A) F , B) T , C) FT , D) $FfFT$. En la siguiente sección buscaremos encontrar motivos repetidos o patrones dentro de estas secuencias.

5.3. Combinación en secuencias

En la sección anterior representamos a las vocalizaciones en un espacio de dos propiedades e identificamos las regiones F , f y T . Una vocalización, entonces, corresponde a una excursión por estas regiones. Pero, ¿Son estos paseos aleatorios? ¿O hay alguna estructura en la manera en que se combinan los distintos fragmentos F , f y T ? Dicho de otra forma, ¿Cómo se comparan las secuencias que aparecen en las vocalizaciones con las que surgirían de combinar dichos elementos al azar?

Para responder a esta pregunta, haremos uso nuevamente de los procesos de Markov (96), como ya lo hicimos en el capítulo 3. Recordemos brevemente que un modelo de Markov consiste en estados y probabilidades de transiciones entre esos estados. Las probabilidades de transición, a su vez, pueden tener memoria, es decir, pueden depender de la historia de los estados anteriores que hayamos visitado. Si el sistema no tiene ningún tipo de memoria, diremos que se trata de una cadena de Markov de orden 0; si depende del estado actual será de orden 1; si depende del actual y el anterior será de orden 2 y así siguiendo.

Dado que dijimos que queremos comparar las cadenas de secuencias que producen las ratas con cadenas al azar, sería natural compararlas con un modelo de orden 0 donde la probabilidad de saltar, por ejemplo, a un estado F dependa únicamente de la proporción total de segmentos F que la rata haya producido. Sin embargo, esto sería automáticamente distinto de las secuencias que conseguimos en la sección anterior dado que mientras la vocalización permanezca en una región del espacio de propiedades, por ejemplo la región F , se la identifica con una única letra. Es decir que nunca veremos una secuencia con dos letras iguales seguidas, a pesar de que en un modelo de orden 0 son inevitables. Para poder evitar las repeticiones hay que tener alguna memoria, al menos del estado actual. Usaremos, entonces, un modelo de orden 1. Para tener en cuenta que las secuencias tienen duración finita, introduciremos un estado más (además de T , F y f) que representará una separación S (un espacio, si estuviéramos pensando

en palabras). Las probabilidades de transición serán independientes del estado, con la salvedad de que no se puede repetir letra. Los valores de las probabilidades estarán dados por la cantidad de S, F, f y T que aparezcan en los datos. Presentamos una representación esquemática de la idea para la construcción del modelo de Markov en la figura 5.11.

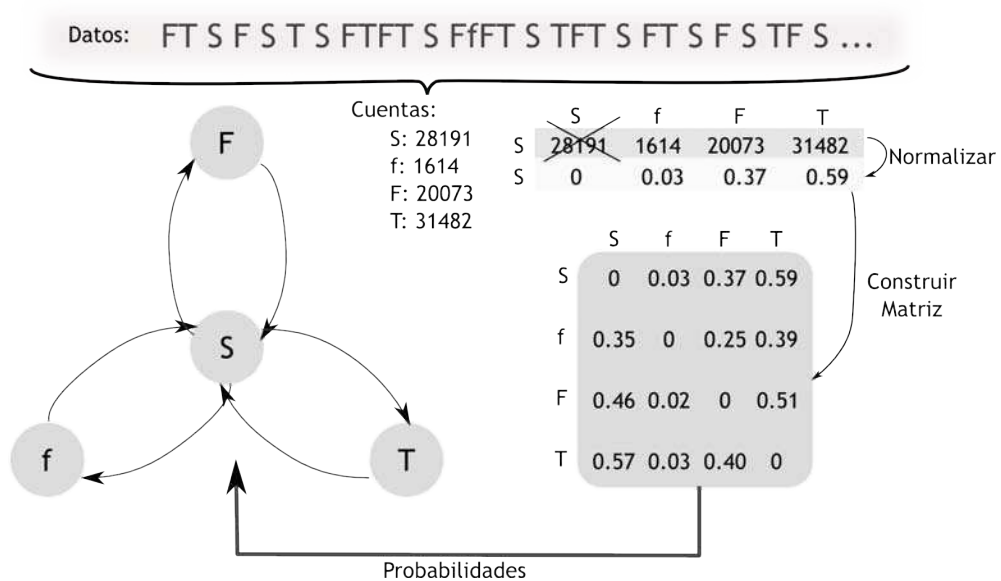


Figura 5.11: Modelo de Markov - Esquema para la construcción del modelo de Markov de orden 1. Tres estados corresponden a tipos de segmentos de vocalización (F,f,T) y uno a una separación o espacio (S). Las probabilidades de transición se estiman a partir de los datos. Se cuenta cuantos elementos hay de cada tipo. Luego se construyen las transiciones para cada uno de ellos simplemente colocando las cuentas en fila y poniendo los elementos diagonales a cero (el ejemplo se muestra para transiciones partiendo desde S). Finalmente, se normalizan las filas para que sumen uno.

Ahora sí, podemos comparar las secuencias que genera el modelo con las que obtuvimos de las grabaciones de las ratas. Para eso, podemos contar que proporción del total representa cada secuencia ya sea generada por el modelo u observada en los datos. Presentamos ese gráfico en la figura 5.12. Ahí se puede ver que hay varias secuencias que aparecen sobre o subrepresentadas en los datos, en comparación al modelo aleatorio. Las distribuciones son claramente distintas (nótese la escala de los intervalos de confianza para la media), es decir, las secuencias no se combinan aleatoriamente de acuerdo a nuestro modelo de Markov. Aparecen ciertas asimetrías interesantes. Consideremos, por ejemplo el par FT y TF. Como contienen las mismas letras, el modelo les asigna

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

idéntica probabilidad. Sin embargo, en los datos se observa 13 veces más la diada FT que TF. Más aún, FT aparece tres veces más de lo que esperaríamos por azar mientras que TF aparece solo con un tercio de la probabilidad que le asigna el modelo aleatorio. Esto indica que la categoría *Flat-Trill combination* definida por Wright está en realidad compuesta en su mayoría por combinaciones del tipo FT. También se observan otras secuencias que aparecen más de lo que esperaríamos por azar. Si filtramos todas las secuencias pidiendo que aparezcan incrementadas en más de 3σ respecto del valor aleatorio obtenemos la siguiente lista: T, FT, TFT, fFT, FfFT y FfT. Es llamativo que los *flats* aparecen solos menos de lo que esperaríamos por azar.

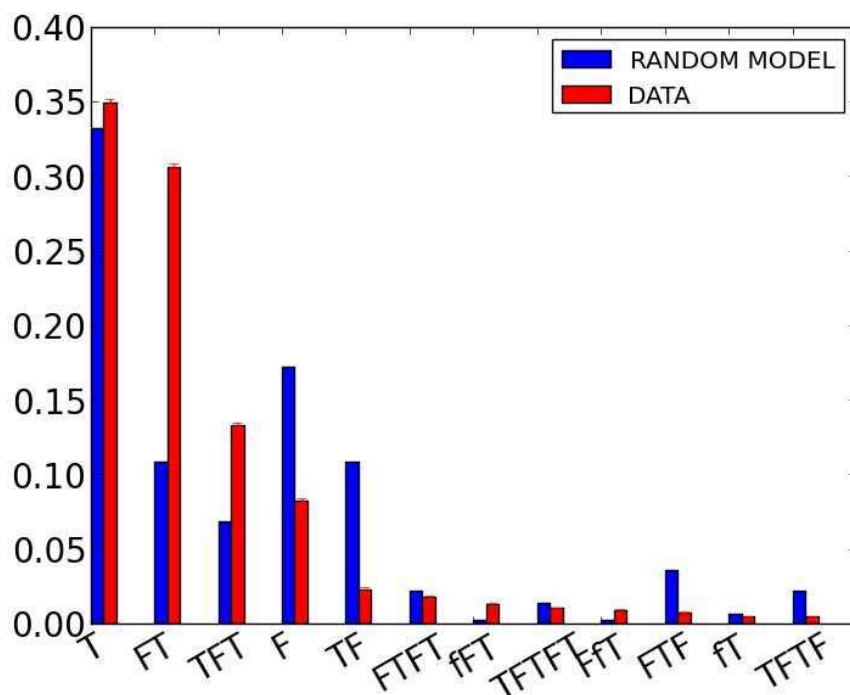


Figura 5.12: Comparación de secuencias con el modelo - En barras rojas y en orden descendente mostramos las 12 secuencias más observadas en los datos. En barras azules mostramos las probabilidades asignadas por el modelo aleatorio a cada una de esas secuencias. Se observan claras discrepancias entre ambas distribuciones.

En lo que resta de esta sección intentaremos ver como estas secuencias se relacionan con las categorías definidas por Wright. Para eso, categorizaremos manualmente 700 vocalizaciones (100 por cada una de las secuencias sobrerrepresentadas y agregaremos 100 también para incluir a los F puros). El resultado de esta categorización puede verse

en la figura 5.13. Las categorías automáticas F y T coinciden con las categorías a ojo en un 88 % y un 79 % respectivamente. Los porcentajes restantes se encuentran divididos en las demás categorías. En el caso de las secuencias FfT corresponden en un 99 % a la categoría *split* mientras que la secuencia FFFt corresponde en un 91 % a *split*. En el caso de las secuencias FT, se reparte casi en partes iguales en *flat-trill* (41 %) y *flat* (45 %). Esto se debe a que algunos *flats* presentan una modulación rápida en frecuencia hacia abajo al final de la vocalización; arrojando valores de la regresión compatibles con la región T. En la figura 5.14 mostramos un ejemplo de esos *flats* junto con la región donde el valor de la modulación entra en la región T. Es llamativo que la secuencia fFT corresponde mayormente a la categoría *flat-trill* (74 %).

En principio uno podría pensar que se trata de un artefacto y que la secuencia fFT corresponde simplemente a un FT que empieza un poco más abajo en frecuencia, tanto así que cruzan a la región f. En ese caso la discretización en dos categorías sería artificial. Sin embargo, el hecho de que la secuencia fFT no se confunda casi nunca con un *flat* nos da la pauta de que quizá sean efectivamente dos cosas distintas. Podemos mirar entonces la distribución de frecuencias con las que comienzan las F, FT y fFT. Si tomamos el primer milisegundo de cada una de esas secuencias y hacemos un histograma de frecuencias obtenemos el gráfico de la figura 5.15. En el panel de arriba se observa como la distribución de fFT se encuentra corrida hacia frecuencias más bajas y es prácticamente disjunta de las distribuciones tanto de FT como de F. Dado que dentro de las secuencias FT vimos que hay una cierta “contaminación” de *flats* podría argumentarse que la distribución de FT y F se parecen artificialmente. Por esa razón, mostramos en el panel de abajo de la misma figura las distribuciones dentro de la categoría FT tanto para las que fueron categorizadas a ojo como *flat* como las *flat-trill*. Se puede ver que se parecen a las distribuciones de FT y F pero no a la de fFT. Todo esto parece indicar que la secuencia fFT es, en efecto, una categoría en si misma distinta de la categoría FT. La transición de f a F, sin embargo, no es con un salto, sino continua.

Finalmente, la secuencia TFT no parece corresponderse fuertemente con ninguna de las categorías de Wright, aunque es interesante que es la que presenta mayor componente de Multi-step (28 %).

En resumen, en esta sección estudiamos algunas propiedades sonoras de las vocalizaciones y vimos que se distribuyen en forma discreta, de manera tal que es posible

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

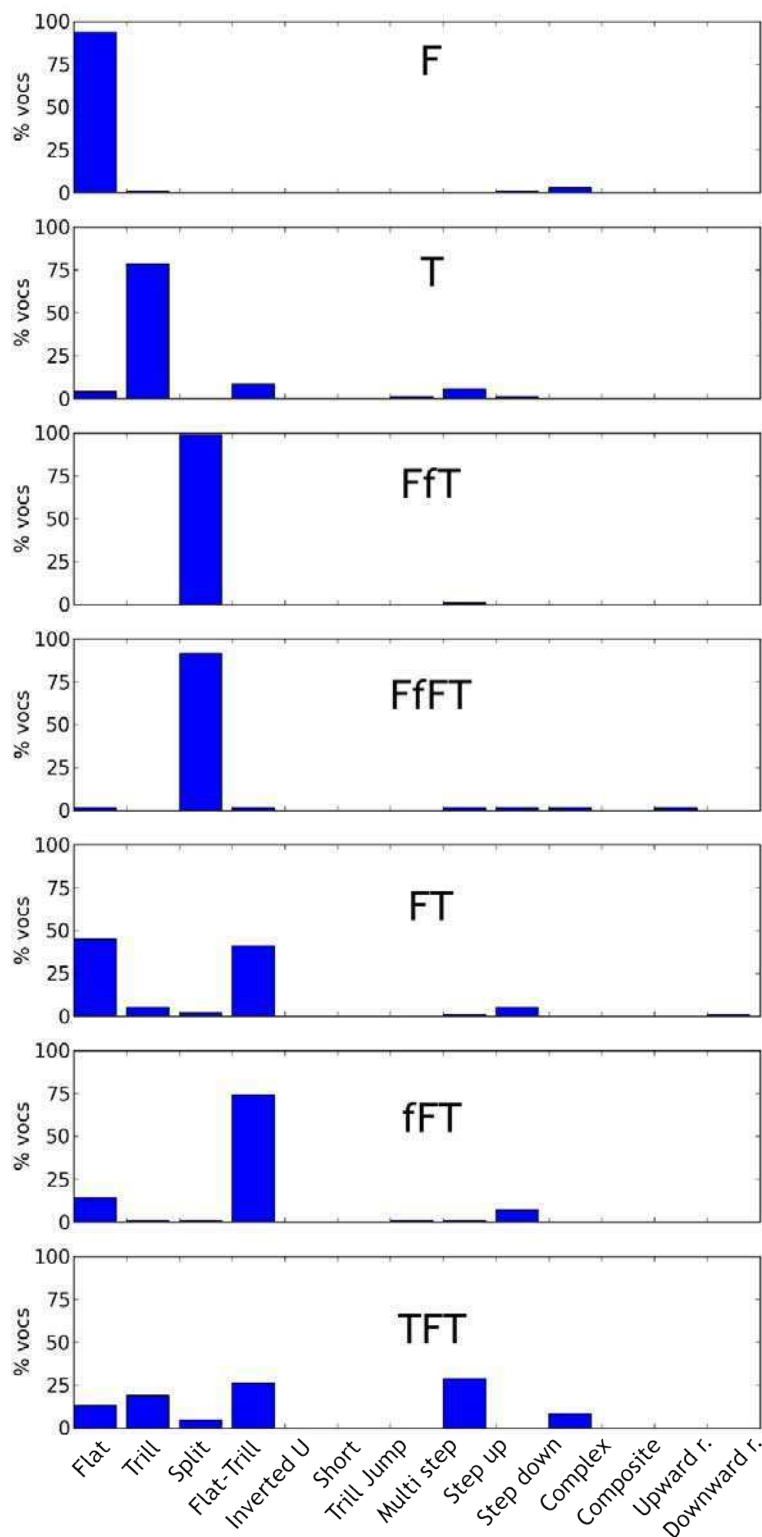


Figura 5.13: Comparación de clasificaciones - Vocalizaciones categorizadas a mano siguiendo los criterios de Wright separadas de acuerdo a la clasificación automática por secuencias sobrerrepresentadas.

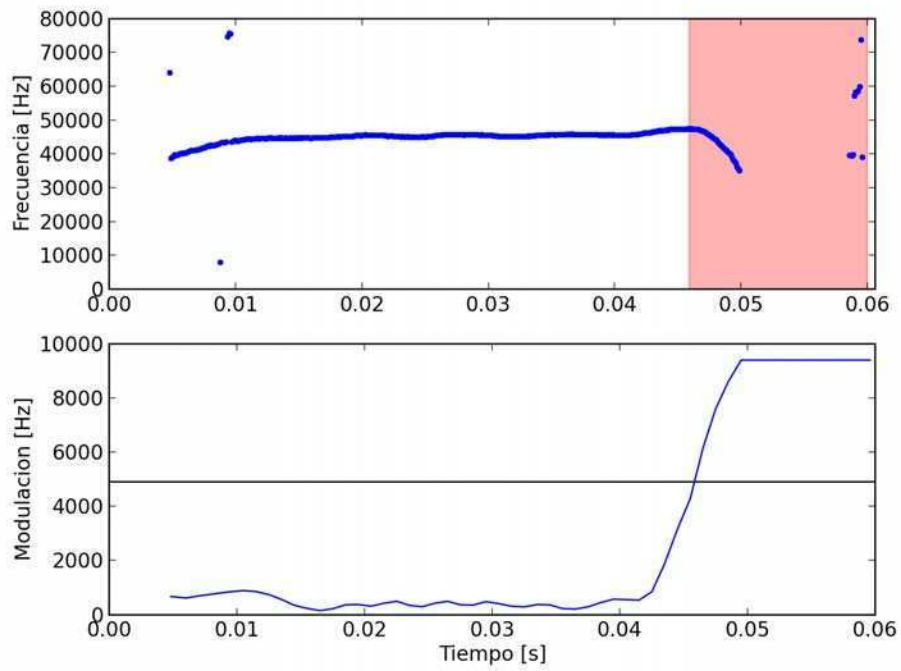


Figura 5.14: Ejemplo de un FT mal categorizado - Se puede ver como en la parte final presenta una modulación que es capturada por los parámetros de la regresión

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

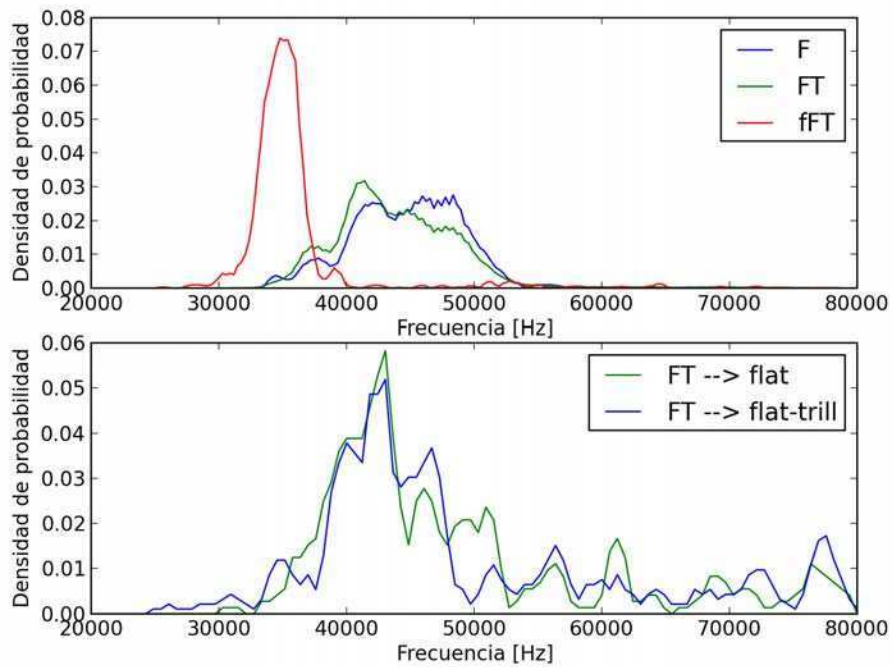


Figura 5.15: Distribución de frecuencias iniciales según secuencias - En el panel superior se muestra la distribución de frecuencias al inicio de secuencias F, FT y fFT. Se puede ver que las fFT tienen una distribución distinta tanto a F como a FT. En el panel de abajo se muestran las distribución para las FT que fueron categorizada a mano separadas según si fueron asignadas a la categoría *flat* o a *flat-trill*.

distinguir tres regiones distintas en el espacio de propiedades (f, F y T). Vimos también que una vocalización explora esas regiones de manera no aleatoria y pudimos reconocer algunos patrones o secuencias en la forma de recorrer esas regiones. A partir de todo esto, nos fue posible reobtener algunas de las categorías descritas en el trabajo de Wright pero que en lugar de estar definidas a ojo, éstas emergieron naturalmente de observar los patrones espectral-temporales de las vocalizaciones ultrasónicas. Sin embargo, no observamos todas las categorías definidas en ese trabajo y surge la duda de si son efectivamente categorías o son particiones arbitrarias dentro de un continuo. Por otro lado, vimos que la categoría *flat-trill combination* que para Wright es cualquier combinación de *flats* y *trills* en realidad está mayormente formada por secuencias del tipo FT y que existe una subcategoría que comienza de frecuencias más bajas la fFT.

Definidas las clases de vocalizaciones y la forma de clasificarlas automáticamente surge la pregunta natural acerca de cómo estas clases de vocalizaciones se estructuran en el tiempo. A esto nos abocaremos en la siguiente sección.

5.4. Estructura entre vocalizaciones

Al estudiar como se estructuran las categorías de vocalizaciones en el tiempo surge una primera pregunta análoga a la que ya surgió para decidir que es una vocalización: ¿Cómo segmentar? En el caso de las vocalizaciones vimos que había una unidad natural para segmentarla dada por el ritmo de respiración. Para segmentar vocalizaciones podemos recordar algo que describimos en el capítulo anterior y es que las vocalizaciones vienen en ráfagas (*bouts*) de aproximadamente 7 Hz que corresponden a momentos en los que todas (o casi todas) las exhalaciones contienen una vocalización. Diremos entonces que si hay un silencio de duración mayor al tiempo típico de dos respiraciones consecutivas, se ha interrumpido la ráfaga. Estudiemos, entonces, la composición de estas ráfagas.

5.4.1. Composición de categorías en una ráfaga

Lo primero que vamos a estudiar es como varía la composición de una ráfaga en función de su largo. Dado que miraremos proporciones de cada categoría y para no trabajar con muchas categorías a la vez agruparemos algunas de las secuencias de la sección anterior. A las secuencias FT y fFT las agruparemos y las llamaremos G, a las

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

secuencias TFF' las llamaremos M y a las secuencias Fff'T y FfT las agruparemos y llamaremos S; al resto de las secuencias las identificaremos con ?. De esta manera, cada vocalización queda etiquetada por un único caracter. Un resumen con las equivalencias y curvas de frecuencia típicas puede verse en la figura 5.16. Ahora sí, podemos graficar la fracción de cada una de estas categorías en función del largo de la ráfaga. Este es el gráfico que presentamos en la figura 5.17. Presentamos allí también en esa misma figura la potencia sonora promedio de cada vocalización también en función del largo de la ráfaga. Esto lo hacemos suponiendo que ráfagas cortas y largas pueden corresponder a estados emocionales diferentes y que este cambio puede verse reflejado también en el volumen. En efecto, como se ve en la figura, la potencia sonora aumenta con el largo de la ráfaga. Cabe aclarar que ver un cambio de volumen en el audio registrado no necesariamente implica un cambio de intensidad por parte de la rata dado que los micrófonos usados son muy direccionales. Es decir que cambios de volumen pueden indicar cambios en la dirección de la cabeza. Al estudiar la composición de categorías de las ráfagas, vemos que tienen diferente composición dependiendo de su largo. En particular, G y S aumentan su proporción a medida que el largo aumenta mientras que F y T la disminuyen. Recordemos que la categoría G esta compuesta de FT y fFT. En ese sentido, es llamativo que F y T se comportan de manera opuesta cuando están combinados en una única vocalización.

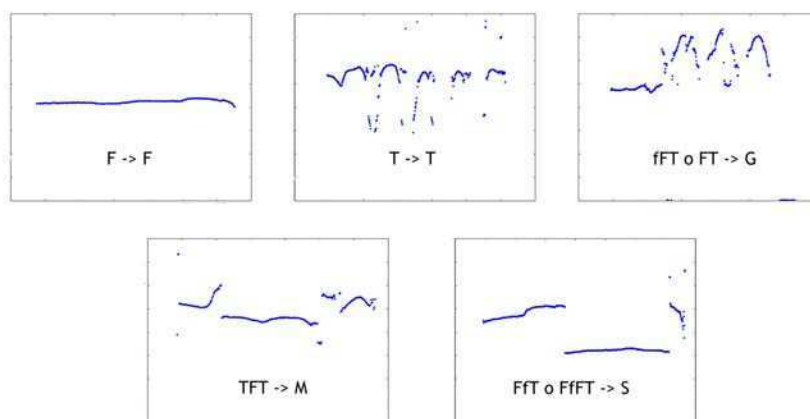


Figura 5.16: Tabla de nueva notación - Nombres de las nuevas categorías en término de las secuencias. Se muestran con una curva de frecuencias representativa de esa categoría.

Ahora bien, como estamos mirando un promedio entre varios animales, esto podría

5.4 Estructura entre vocalizaciones

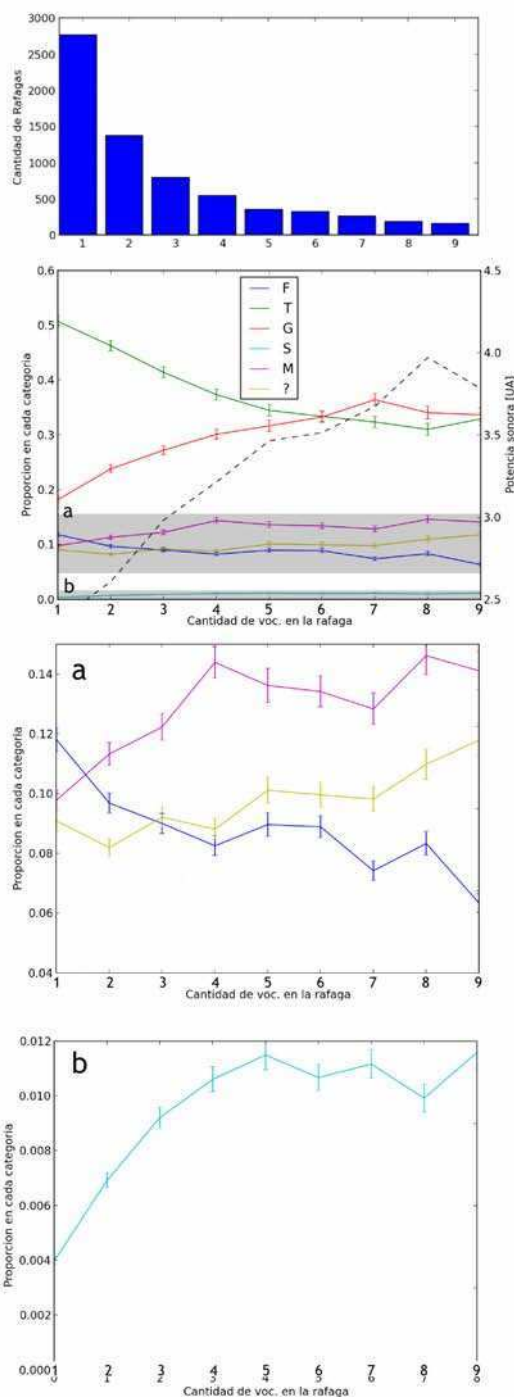


Figura 5.17: Composición de las ráfagas en función de su largo - El panel superior muestra la cantidad de ráfagas en los datos según su largo. Los dos paneles inferiores son zonas ampliadas del segundo superior. Se puede ver como varía la composición de las categorías en función del largo de la ráfaga. En particular se ve un claro aumento de las categorías G y S y una disminución de las categorías F y T. En la escala de la derecha del segundo panel se presenta también la potencia sonora promedio de las vocalizaciones en función del largo de la ráfaga

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

deberse, no a que la rata cambia la composición según el largo de la ráfaga sino a que hay diferencias entre ratas. Podría pasar que haya ratas que hacen ráfagas más cortas y que a su vez emiten más vocalizaciones T, por ejemplo. Sin embargo, esto no es así. Si repetimos el análisis de la figura 5.17 vemos que se observa el mismo patrón para todas ratas estudiadas, cada una por separado. En la figura 5.18 mostramos esto para las categorías T y G en todas las ratas excepto en V_1 que posee muy pocas vocalizaciones y solo presenta un patrón ruidoso. Se evitan las barras de error para no oscurecer la visibilidad de la figura.

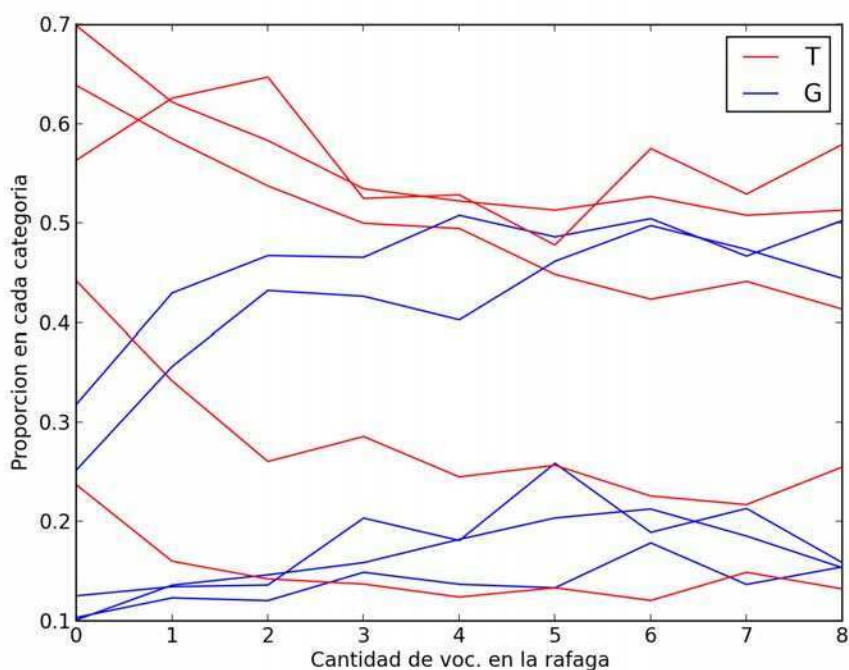


Figura 5.18: Composición de la ráfaga en función de su largo (varias ratas) - Mostramos las proporciones de T y G que presentan un patrón similar en todas las ratas T disminuye y G aumenta. No se muestran las barras de error para no oscurecer la visibilidad del gráfico. No se incluyen los datos de V_1 que emitió muy pocas vocalizaciones y presenta solo un comportamiento ruidoso.

El aumento de volumen con el largo de la ráfaga apoya la idea de que lo que estamos observando es un cambio en el estado de excitación de la rata. Bajo esta hipótesis, las categorías G y S denotarían estados de mayor excitación mientras que F y T señalarían una excitación más baja y no tendría por qué haber un orden dentro de la ráfaga.

Alternativamente, podría ser que los cambios en densidad de cada categoría se deban, de hecho, a un efecto de orden. Por ejemplo, si todas las ráfagas comenzaran con mayor probabilidad con Ts pero luego tuvieran con mayor probabilidad Gs también veríamos un efecto con la longitud de la ráfaga. Esto es, efectivamente, lo que ocurre. Para sacarnos de encima el efecto con la longitud podemos mirar únicamente secuencias de un dado largo. En la figura 5.19 presentamos tanto las proporciones de cada categoría como la potencia sonora en función del orden dentro de cada ráfaga. Los promedios están separados en paneles por ráfagas de igual largo. Se puede observar que los primeros lugares de cada ráfaga tiene mayor probabilidad de T mientras que los últimos tienen mayor probabilidad de G. Es decir, hay cierto orden dentro de una ráfaga. Hay un pequeño efecto del volumen con el orden, y un claro efecto con el largo de la ráfaga (el valor promedio del volumen sube a medida que avanzamos de panel).

En resumen, observamos que las categorías definidas en este capítulo no se emiten homogéneamente en el tiempo sino que tienen un orden. Y que ese orden parece estar atado a una cadencia natural de emisión de vocalizaciones en ráfagas. La emisión parece ser, en esencia, probabilística, lo cual la distingue de otras formas de vocalización más estereotipadas (como puede ser por ejemplo el canto de aves). Ahora bien, que no parezca seguir siempre una misma regla y que sea probabilística no quiere decir que sea simplemente al azar. De hecho, el efecto de orden que observamos en esta sección nos dice que las probabilidades de emisión no pueden ser independientes entre sí. En el siguiente apartado veremos cual es el grado de dependencia temporal que existe, es decir, cuántos pasos hacia atrás hacen falta para hacer una buena predicción del sistema.

5.4.2. Memoria del sistema

En esta sección haremos nuevamente uso de los procesos de Markov pero ahora los estados serán los distintos tipos de vocalizaciones que fuimos definiendo: F, T, G, M, S y ?. Vimos que las probabilidades de transición del paso n en un modelo de Markov pueden depender, en general, de los estados que el sistema haya recorrido en los pasos $n-1$, $n-2$, ..., $n-l$. Se suele decir que el sistema tiene una memoria de l pasos. Sin lugar a dudas, a partir de una secuencia finita de datos observados, aumentar el orden del modelo hará que los datos sean mejor y mejor ajustados por el modelo. Dicho de otra forma, el máximo de la función verosimilitud aumentará a medida que aumentemos el orden

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

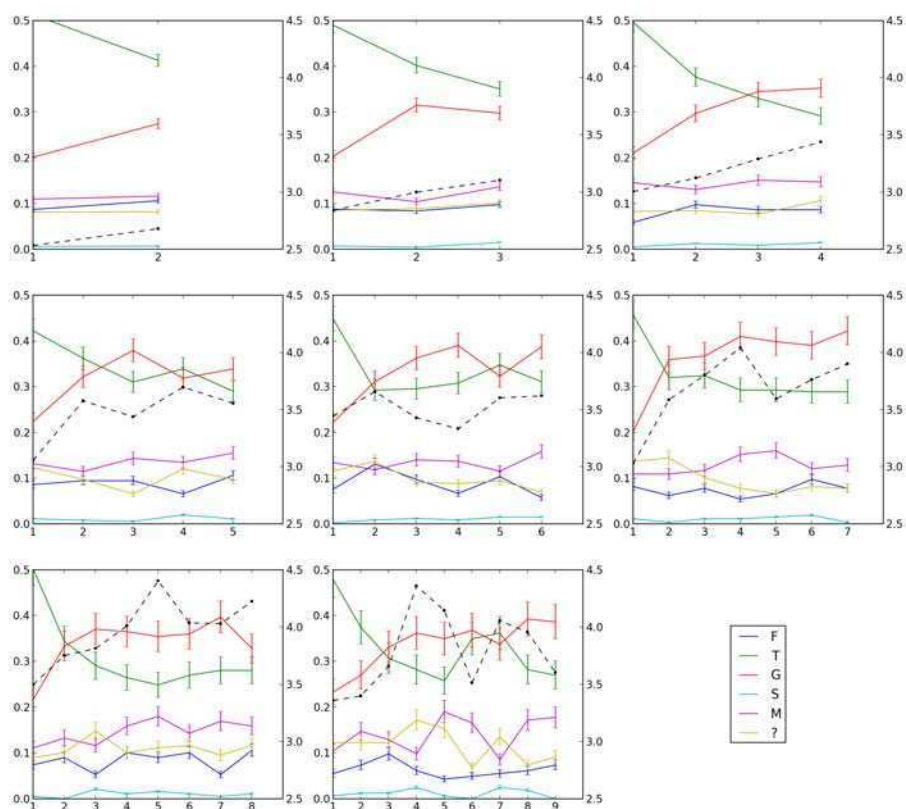


Figura 5.19: Proporción de Categorías en función del orden - Cada panel corresponde a promedios para ráfagas de largo creciente. El panel de arriba a la izquierda corresponde a ráfagas que contienen solo dos vocalizaciones mientras que el de abajo a la derecha corresponde a ráfagas de largo nueve. En el segundo eje de cada gráfico muestra la potencia sonora promedio correspondiente a cada posición. Se puede observar que los primeros lugares de cada ráfaga tiene mayor probabilidad de T mientras que los últimos tienen mayor probabilidad de G. Es decir, hay cierto orden dentro de una ráfaga. Sin embargo no hay ninguna tendencia clara del volumen con el orden mientras que si se observa un efecto con el largo de la ráfaga.

del modelo. Recordemos que la función verosimilitud es simplemente la distribución de probabilidad de los datos observados como función de los parámetros del modelo. Por ejemplo, en el caso de una moneda, si asumimos que el proceso es binomial y llamamos a la probabilidad de cara p , la función verosimilitud de la secuencia observada: XOXXOXOOOO donde hay 6 caras y 4 cecas será $L(p) = p^6(1-p)^4$. Por supuesto que al incluir más y más parámetros en nuestro modelo haremos mejores ajustes de nuestros datos pero, a la vez, caemos en el riesgo de estar sobreajustando, es decir, de haber construido un modelo que solo explique las peculiaridades de la tira de datos particular que estamos analizando. Si queremos comparar la bondad de ajuste de modelos con distinta cantidad de parámetros tenemos que de alguna manera penalizar a los modelos que usan más parámetros. Hay distintas propuestas para hacer esto, en esta sección utilizamos dos: el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC). Ambas propuestas son similares y corresponden a evaluar la bondad del ajuste con algún término que penalice usar un mayor número de parámetros. Las propuestas son las siguientes:

$$\begin{aligned} AIC &= -2\ln(L_{max}) + 2k \\ BIC &= -2\ln(L_{max}) + k\ln(N) \end{aligned} \tag{5.3}$$

donde L_{max} es la máxima verosimilitud, k la cantidad de parámetros del modelo y N la cantidad de datos observados. BIC es más restrictivo que AIC, pues $\ln(N) > 2$ para $N \geq 8$. Para una derivación de cada uno de los métodos pueden referirse al trabajo de Burnham y Anderson (97). Como idea intuitiva del origen del término $k\ln(N)$ en BIC imaginemos lo siguiente. Si un modelo posee k parámetros en un vector $\vec{\theta} = (\theta_1, \dots, \theta_k)$ podemos escribir la verosimilitud de los datos D dada la hipótesis H_i como:

$$P(D|H_i) = \int P(D|\vec{\theta}, H_i)P(\vec{\theta}|H_i)d\vec{\theta} \tag{5.4}$$

Asumiendo que la distribución de los parámetros se puede aproximar por un pico, la integral queda aproximada por:

$$P(D|H_i) \approx P(D|\vec{\theta}_{max}, H_i) \times P(\vec{\theta}_{max}|H_i)\sigma_{\theta}^k \tag{5.5}$$

es decir, por el valor de su máximo multiplicado por el volumen de un cubo k -

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

dimensional de lado σ_θ . El primer término es simplemente lo que llamamos L_{max} . El término $P(\vec{\theta}_{max}|H_i)$ corresponde al prior que elijamos para cada parámetro y que si no tenemos ninguna información previa podemos tomarlo como una constante. σ_θ corresponde al error en la estimación de cada parámetro que irá disminuyendo a medida que consigamos más datos, es decir que aumente N . Bajo una aproximación Gaussiana $\sigma_\theta \propto \frac{1}{\sqrt{N}}$. Reemplazando todo esto, la ecuación 5.5 queda:

$$P(D|H_i) \approx L_{max} \left(\frac{C}{N}\right)^{\frac{k}{2}} \quad (5.6)$$

y tomando el logaritmo y multiplicando por -2:

$$-2\ln(P(D|H_i)) \approx -2\ln(L_{max}) + k\ln(N) - k\ln(C) \quad (5.7)$$

Finalmente, despreciamos el último término para valores grandes de N y llegamos a la ecuación 5.3.

Ahora sí, estamos en condiciones de evaluar nuestros datos para estimar el orden del modelo de Markov conformado por las distintas categorías, es decir de ver cuál es el rango de dependencia entre distintas vocalizaciones. Construimos, entonces, un modelo de Markov con los estados $[F, T, G, M, S, ?]$ y buscamos el orden del modelo que minimiza los criterios AIC o BIC. Analizando 28201 vocalizaciones, el método AIC arroja un mínimo para un orden de modelo 3 y BIC para un orden de modelo 2. Para estimar cual es la relación que tienen con el siguiente modelo más probable podemos restarle al valor de AIC_{min} (o BIC_{min}) el valor de AIC (o BIC) para el modelo con valor de AIC (BIC) más cercano, en este caso orden 2 (orden 1 para BIC). Esto arroja un $\Delta AIC = -42.8$ y un $\Delta BIC = -61.3$. Ambas son diferencias significativas; dado que son escalas logarítmicas esas diferencias corresponden a cocientes de verosimilitudes de 10^{-10} y 10^{-14} respectivamente. Si bien ambos criterios no coinciden, los dos muestran que el sistema tiene memoria, es decir, un modelo de orden cero no es una buena explicación de los datos, más aún, el más conservador de esos criterios estima una memoria del sistema de orden 2. Es decir, que la probabilidad de emisión de cada vocalización depende de cuales sean las dos vocalizaciones que se emitieron antes. Esto no nos garantiza que existen reglas para la construcción de secuencias de vocalizaciones pero indica la existencia de estructura en la forma en que se organizan temporalmente.

5.5. Interacción entre individuos

Como última parte de este capítulo, exploramos las interacciones vocales entre individuos en la arena de diálogos. Analizamos las distribuciones de tiempos entre vocalizaciones separándolos en dos grupos: tiempos intra-rata (es decir tanto la vocalización previa como la posterior a ese intervalo corresponden a la misma rata) y tiempos inter-rata (tiempos entre vocalizaciones de distintas ratas). Ambas distribuciones pueden verse en la figura 5.20. Para estudiar, en qué medida la distribución de tiempos INTER está dada por la distribución de tiempos INTRA podemos generar datos subrogados que permuten aleatoriamente los intervalos de tiempo de manera de preservar las distribuciones INTRA pero destruir toda estructura en la distribución INTER. Eso es lo que se muestra también en la figura 5.20. No se observan diferencias apreciables entre la distribución INTER de los datos reales y los datos subrogados, a excepción de un pequeño pico cercano al cero. Sin embargo, creemos que ese pico puede deberse a un artefacto de nuestro *setup* a la hora de asignar el tiempo a dos vocalizaciones que ocurren casi simultáneamente. Dado que contamos con datos para interacciones de otro experimento con ratas machos y ratas hembra, repetimos el mismo análisis para ese juego de datos y es lo que presentamos también en la figura 5.20 pero en el panel de la derecha. En esos datos sí se observa una diferencia significativa entre los datos reales y los subrogados para tiempos entre los 0 y los 100 ms. Se observa una reducción de aproximadamente 20 % respecto a la cantidad de vocalizaciones que se esperarían por azar. Nuevamente aparece también un pico cercano a tiempo cero que creemos artefactual. Una diferencia a notar entre los grupos de ratas de ambos experimentos es que en el experimento de machos las ratas utilizadas se usaron para poner a punto el dispositivo experimental y por lo tanto, al momento de tomar los datos ya habían tenido varios encuentros en la arena de diálogo. En cambio, en el experimento de machos y hembras, los sujetos no habían tenido encuentros previos.

Si bien el efecto puede parecer sutil, es la primera evidencia de algún tipo de interacción vocal entre individuos en la rata. Es una muestra de que son capaces de procesar las vocalizaciones en una escala temporal rápida y modificar su comportamiento (vocalizar menos).

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

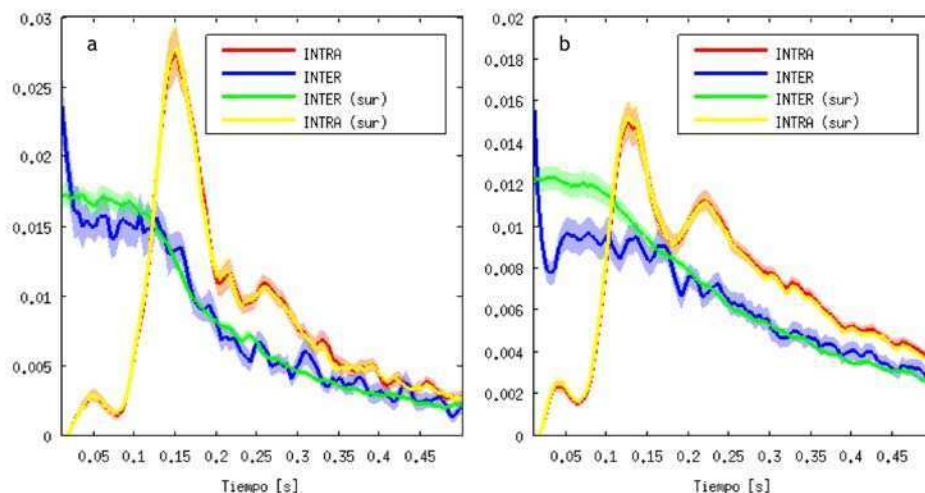


Figura 5.20: Histogramas de Tiempo entre vocalizaciones para dos experimentos - Panel (a): Experimento de machos V1-V6, Panel(b): experimentos de machos y hembras. Los cuatro colores corresponden a los tiempos entre dos vocalizaciones consecutivas de: Azul) una misma rata (INTRA), Rojo) vocalizaciones de dos ratas distintas (INTER). Verde y amarillo corresponden a permutaciones aleatorias que preservan la distribución INTRA pero rompen toda estructura en INTER.

5.6. Conclusiones

En este capítulo hicimos una suerte de anatomía de las vocalizaciones buscando en ellas una jerarquía de estructuras. Empezamos definiendo propiedades sonoras para cada pequeño intervalo de tiempo y vimos que esas propiedades se agolpan en nubes de ciertos valores. Luego, estudiamos las transiciones que se dan entre esas nubes de propiedades y encontramos secuencias estereotipadas que nos sirvieron para definir clases de vocalizaciones. Pudimos relacionar y contrastar nuestras categorías, que emergieron de la estructura en los datos, con la única otra categorización existente (definida ad-hoc por el grupo de Wright). Finalmente, en el nivel superior estudiamos como estas categorías se estructuran en el tiempo y observamos que la escala natural para analizarlas era la de las *ráfagas*. Así, observamos que hay un efecto de orden dentro de cada ráfaga y que las distintas categorías no se distribuyen uniformemente. Logramos estimar cual es la memoria del sistema, es decir cuántas vocalizaciones anteriores determinan la próxima vocalización (en un sentido Markoviano). Usando el criterio más conservador el sistema tiene memoria de orden 2 y con un criterio más laxo es de orden 3. Finalmente, estudiando las interacciones entre individuos, logramos observar (en algunos de

5.6 Conclusiones

los experimentos) que la vocalización de una de las ratas inhibe temporariamente y de manera parcial las vocalizaciones de su compañera.

5. CLASIFICACIÓN AUTOMÁTICA Y ESTRUCTURA DE VOCALIZACIONES

6

Otros aspectos de la producción vocal en ratas

6.1. Introducción

Dado que el campo de las vocalizaciones en roedores está relativamente subexplorado, los datos obtenidos en la arena de diálogos permiten estudiar muchas preguntas más que la de clasificación desarrollada en el capítulo anterior. En este último capítulo presentamos algunos resultados cortos que surgen del análisis de los datos obtenidos en la misma arena de diálogos; son dos pequeñas exploraciones que estudian: 1) La relación entre vocalización y locomoción, 2) Los mecanismos de producción sonora. Así mismo, presentamos también en este capítulo, un último experimento que explora el grado de control voluntario de las ratas en la iniciación de una vocalización.

6.2. Vocalización y locomoción

La relación entre vocalizaciones ultrasónicas y locomoción en roedores no ha sido prácticamente estudiada. Los únicos trabajos, de los que estamos al tanto, que investigan esta relación son el de Thiessen y Kittrell (98) de 1980 y el de Blumberg (99) de 1992. En el primero de esos trabajos, los autores muestran que existe una correlación entre la cantidad de vocalizaciones que emiten estudiando Gerbillos (*Gerbils*) y la cantidad de eventos de movimiento (etiquetados por ellos por inspección visual de video), promediando en la escala del minuto. Blumberg va un poco más allá y sugiere directamente que las vocalizaciones ultrasónicas de los roedores son directamente un

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

subproducto de la locomoción (análogo al rebufar de un caballo cansado). Sin embargo, su observación se basa únicamente en una apreciación cualitativa de videos en cámara lenta.

Aprovechando que tenemos información de video sincrónica con la grabación de audio podemos estudiar la relación entre la posición de la rata y su tasa de emisión de vocalizaciones. Inspeccionando un gráfico de velocidad en función del tiempo en conjunto con los tiempos en los que ocurren las vocalizaciones (ver figura 6.1) podemos observar que hay momentos donde vocalización y locomoción se dan simultáneamente.

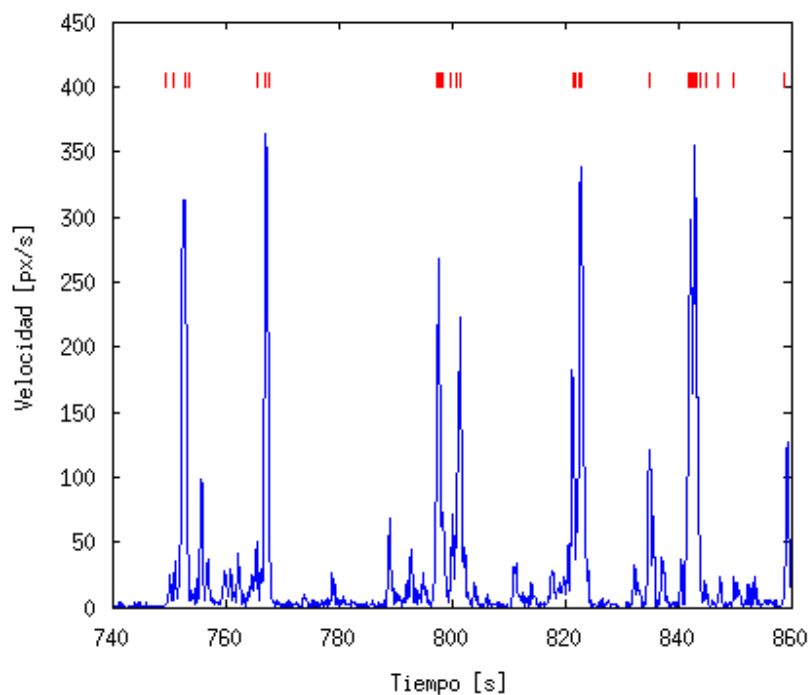


Figura 6.1: Velocidad y vocalizaciones - Ejemplo de un registro de velocidad donde se observa una gran sincronía entre momentos de alta velocidad y vocalizaciones (indicadas por líneas verticales en la parte de arriba)

Para estudiar esto de manera cuantitativa podemos construir un filtro lineal que integre la información de comportamiento vocal de manera tal de reproducir, lo mejor posible, la curva de velocidades. Es decir si V_i es la velocidad para cada cuadro i de video, definimos R_i un vector que vale 1 si el i -ésimo cuadro contiene una vocalización y 0 si no. Luego definimos el filtro W_k que nos permitirá generar un vector \hat{V}_i de

velocidades estimadas a partir de la convolución de \mathbf{R} y \mathbf{W} . Es decir,

$$\hat{V}_i = \sum_{j=-K}^K W_j R_{i+j} \quad (6.1)$$

Así escrito, se trata de un filtro no-causal pues para calcular el punto i -ésimo de \hat{V} se usan K puntos que están en el futuro y K puntos del pasado. Una estimación del filtro \mathbf{W} se puede obtener mediante cuadrados mínimos pidiendo que sus coeficientes minimicen la diferencia entre \mathbf{V} y \hat{V} .

Haciendo esto obtenemos la figura 6.2. Ahí mostramos en el panel (a) los filtros obtenidos para las distintas ratas y sesiones, es decir los valores de W_j ; en el panel (b) los valores de correlación entre \mathbf{V} y \hat{V} también para cada sesión y rata en función del orden K del filtro. En los paneles inmediatamente inferiores mostramos la mediana de esas magnitudes entre sesiones y ratas. Podemos observar que el filtro no se encuentra centrado respecto del cero; esto quiere decir que las vocalizaciones preceden levemente a la curva de velocidad. El retraso promedio es de 5 cuadros de video y dado que las cámaras grababan a, aproximadamente, 15 Hz el retraso en tiempo sería de un tercio de segundo. Por otro lado, el tamaño de la ventana en que el filtro integra, es de aproximadamente 20 cuadros. Es decir que el filtro integra en una ventana de 1,3 segundos. Para calcular el valor de la correlación que se muestra tanto en los paneles (b) como (d) calculamos los valores del filtro usando la mitad de los datos de cada sesión y luego calculamos la correlación para la otra mitad de los datos. Ésta es una técnica usual para evitar sobre-ajustar los datos. Si simplemente miráramos la correlación sobre los mismos datos con los que calculamos los coeficientes del filtro, cuantos más coeficientes ajustemos mejor será el resultado. En cambio, si usamos una mitad para entrenar los coeficientes de \mathbf{W} y la otra para medir la bondad del ajuste (en este caso con la correlación) en cuanto empezamos a sobre-ajustar los datos la bondad del ajuste disminuirá. Esto es lo que se observa parcialmente en el panel (d). Finalmente, en el panel (e) mostramos un histograma de las mejores correlaciones que se pueden obtener por este método para cada rata y sesión.

En resumen, en esta sección vimos, usando los datos del video sincrónico con las grabaciones de audio, que existe una correlación importante entre la velocidad de la rata y la tasa de emisión de vocalizaciones. Observamos, además, que la emisión de vocalizaciones precede al aumento de la velocidad.

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

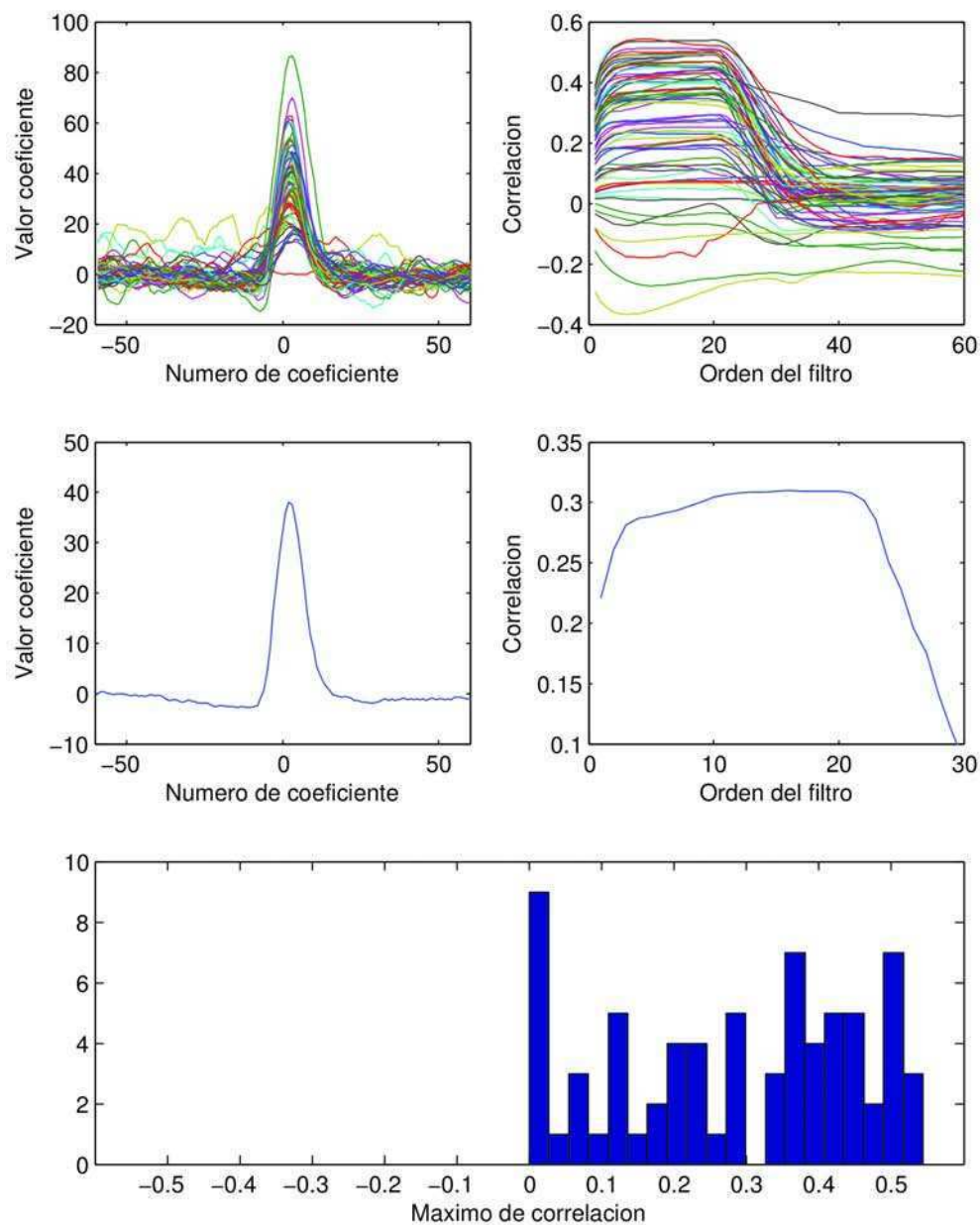


Figura 6.2: Filtros entre Vocalizaciones y velocidad - (a) Coeficientes de un filtro lineal de Wiener entre la secuencia de vocalizaciones y la velocidad (b) correlación en función del orden del filtro para cada rata y cada sesión. (c) y (d) son medianas de los paneles superiores. (e) Tiene un histograma de los valores de correlación en cada sesión.

6.3. Mecanismos de producción

Como discutimos en el capítulo introductorio, los mecanismos de producción de las vocalizaciones ultrasónicas no están del todo claros. En esta sección, aportamos algunas evidencias a favor de modelos aerodinámicos del tipo silbato. Para eso, estudiaremos las características de los saltos en frecuencia que presentan muchas de las vocalizaciones ultrasónicas. Vimos algunos ejemplos de estos saltos en el capítulo anterior (y de hecho fueron un pequeño obstáculo a la hora de definir las propiedades) (ver figura 5.6). Una manera de representar todas las vocalizaciones juntas que evidencia la existencia de saltos es graficar la frecuencia en un dado instante contra la frecuencia en un instante posterior. Es decir, si usamos las curvas de frecuencia en función del tiempo del capítulo anterior y $[F_1 \dots F_N]$ corresponde a las frecuencias en los instantes $[t_1 \dots t_N]$ construiremos un gráfico de F_i versus F_{i+1} . Como se trata de un gráfico con millones de puntos, haremos un histograma que nos permita entender la densidad de puntos en cada región del espacio (F_i, F_{i+1}) . Esto es lo que presentamos en la figura 6.3. La gran mayoría de los puntos cae en la diagonal (es decir que la frecuencia no cambia apreciablemente). Se observan también algunas nubes correspondientes a distintos tipos de salto.

Para entender el origen de los saltos estudiamos el funcionamiento de los silbatos tipo *bird-call* que consisten un orificio por el que pasa un fluido produciendo un chorro que impacta luego en un segundo orificio más adelante. Un esquema puede verse en la figura 6.4. En la misma figura, se ve también como el flujo del chorro se desarrolla convirtiéndose en una serie de vórtices en forma de aro que luego impactan en el segundo orificio produciendo sonido. Parte del sonido regresa al primer orificio y estimula la producción de otros vórtices. Si la velocidad del flujo es U , los vórtices viajan a una velocidad menor $U_c \leq U$; típicamente U_c se encuentra en el rango que va de $\frac{U}{2}$ a U . Asumiendo que hay una dada fase de la onda de sonido a la que se acopla la emisión de vórtices ($\frac{\theta}{2\pi}$) y que la frecuencia de emisión de vórtices es f podemos escribir la condición para que el sistema resuene:

$$2\pi f \left(t_i - \frac{L}{v_0} - \frac{L}{U_c} \right) - 2\pi\theta = 2m\pi \quad (6.2)$$

Donde L es la distancia entre orificios, los $t_i = \frac{1}{f}, \frac{2}{f}, \frac{3}{f} \dots$ son los instantes en los que se emite un vórtice y v_0 es la velocidad del sonido. La condición dice, esencialmente, que la diferencia de fase entre la onda de sonido y la fase θ tiene que ser una cantidad

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

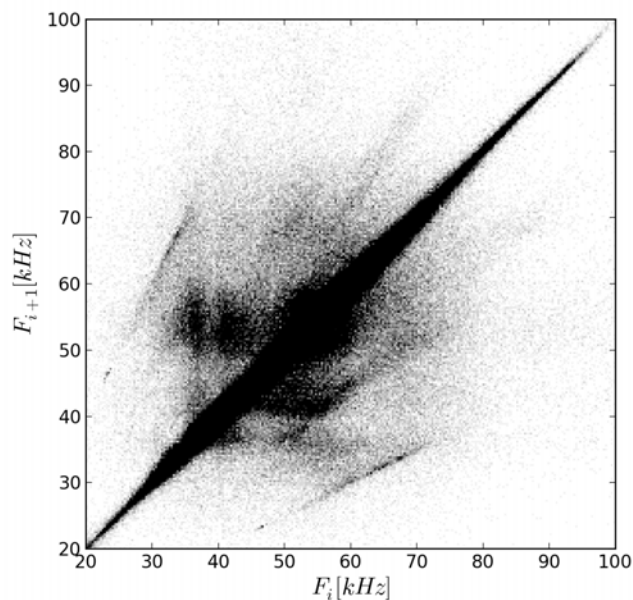


Figura 6.3: Mapa de una frecuencia en un instante contra el siguiente - La gran mayoría de los puntos cae en la diagonal (es decir que la frecuencia no cambia apreciablemente). Se observan también algunas nubes correspondientes a distintos tipos de salto.

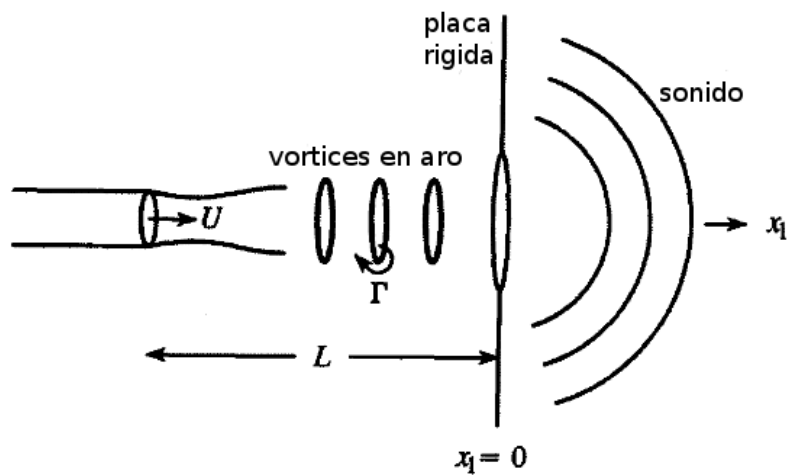


Figura 6.4: Esquema de un silbato *bird-call* - Reproducido del libro "Acoustics of Fluid Structure Interactions" (100)

entera de vueltas. Despejando, y despreciando el término de v_0 por ser $v_0 \gg U_c$, la ecuación queda:

$$f_m \simeq \frac{(m - \theta)U_c}{L} \quad (6.3)$$

Vemos, entonces, que la frecuencia aumenta linealmente con la velocidad U_c y es inversamente proporcional a L . Sin embargo un mismo valor de U_c y de L pueden dar lugar a distintas frecuencias discretas variando el número entero m . Un salto, entonces, corresponde a cambiar de modo. Para comparar la ecuación 6.3 con la figura 6.3 sin conocer ni los valores de U_c ni L pensemos lo siguiente. Sean cuales sean esos valores, deberían ser aproximadamente iguales antes y después de un salto. Por lo tanto, si el salto es entre valores de frecuencia f_m y f_l , el cociente queda:

$$\frac{f_m}{f_l} = \frac{m - \theta}{l - \theta} \quad (6.4)$$

Es decir que las frecuencias antes y después del salto tienen una relación lineal cuya pendiente depende de entre qué dos modos estemos saltando (m y l). Ahora podemos volver a realizar el gráfico de la figura 6.3 agregando las rectas que esperamos de la ecuación 6.4. Esto es lo que presentamos en la figura 6.5. El valor de θ que ajusta las rectas a las nubes es $\theta = 0$. Las rectas dibujadas corresponden a saltos entre modos consecutivos ($1 \rightarrow 2, 2 \rightarrow 3, \dots$). Se pueden ver nubes correspondientes a las pendientes $1/2, 2/3, 3/4$ y $4/5$. Datos experimentales de silbatos tipo *bird-call* indican que $\theta \simeq \frac{1}{4}$. El hecho de que θ valga cero puede indicar que el segundo orificio es, en realidad menor que el primero. En este caso, los vórtices se reflejan y se adquiere una fase extra ψ que puede compensarse con *theta* (ver el libro de Howe (100) pag 486-487).

Habiendo visto como el modelo aerodinámico puede explicar la estructura de saltos resta ver si los tamaños y velocidades que éste requiere son compatibles con la anatomía de la rata. Brudzynski y Fletcher (75) estiman que una velocidad razonable para el flujo de aire en la laringe es de, aproximadamente, 40 m/s; eso daría un rango para U_c de entre 20 m/s y 40 m/s. Lo cual implica que para el modo fundamental ($m = 1$) produzca frecuencias en el rango de 20-40 kHz el valor de L debe ser cercano a 1 mm. En un abstract sin publicar Sanders et al. (54) reportan que, durante la emisión de vocalizaciones ultrasónicas, las cuerdas vocales se encuentran totalmente cerradas y no vibran. Observan también que se forma una pequeña abertura (de aproximadamente

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

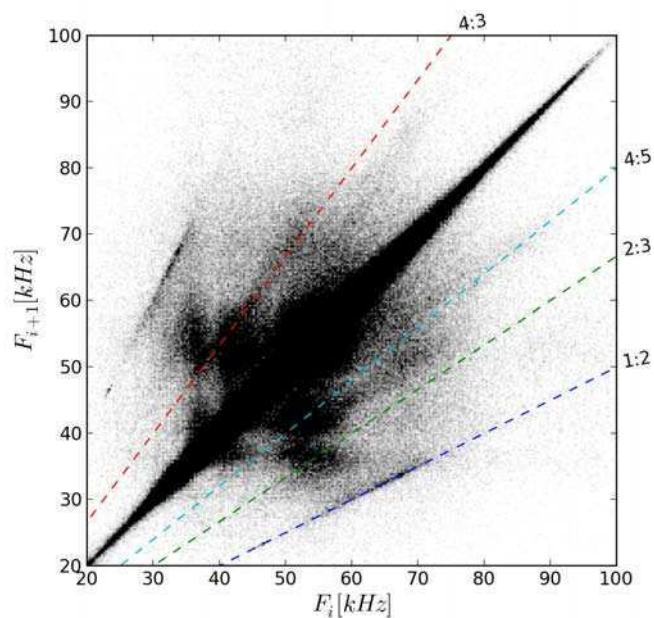


Figura 6.5: Mapa de saltos con rectas dadas por la ec 6.4 - Misma figura que la 6.3 junto con las rectas correspondientes a los distintos saltos. El histograma, en este caso está simetrizado (es decir que se cuentan juntos saltos hacia arriba o hacia abajo). Esto resalta un poco la estructura. La recta de pendiente $3/4$ fue puesta del lado simétrico (como $4/3$) para hacer más sencilla la visualización del gráfico.

1 mm) entre los cartílagos aritenoides. Brudzynski y Fletcher sugieren que el segundo orificio podría estar formado por la oclusión parcial de la epiglotis dando un tamaño de la cavidad de aproximadamente 1 mm (ver figura 6.6).

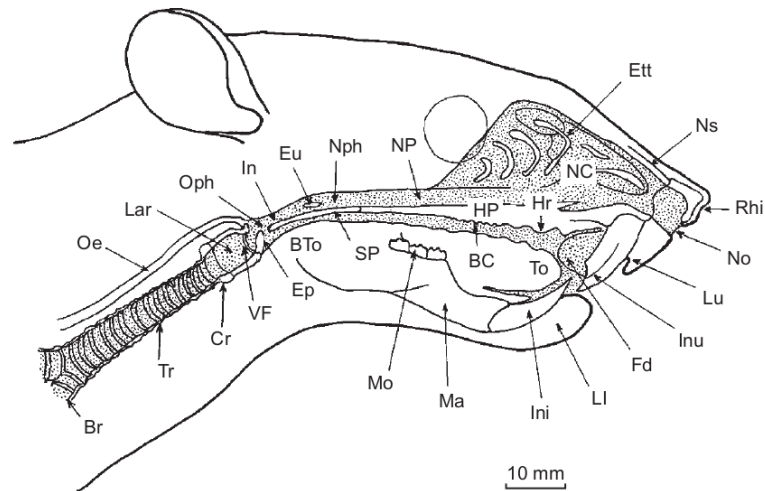


Figura 6.6: Anatomía del tracto respiratorio de la rata - Ep: epiglotis; Ett: huesos etmoides; Eu: apertura de Trompa de Eustaquio; Fd: Pliege cubriendo la diastema; HP: Paladar Duro; Hr: Paladar rugoso; In: Coana; Ini: Incisivo inferior; Inu: Incisivo superior; Lar: Laringe; Ll: Labio inferior; Lu: Labio Superior; Ma: Fragmento inferior Mandíbula; Mo: Tres molares (superiores omitidos); NC: Cavidad Nasal; No: Fosas nasales; NP: Pasaje nasal; Nph: Nasofaringe; Ns: Huesos nasales; Oe: Esófago; Oph: Orofaringe; Rhi: Rinario; SP: Paladar Blando; To: Parte rostral de la lengua; Tr: Tráquea; VF: Cuerdas Vocales. Reproducido de Brudzynski, Fletcher(75)

Un aspecto que no discutimos es que la ecuación 6.3 nos permite calcular qué posibles valores de frecuencias presentará el sistema **en caso de que se desarrolle un flujo de vórtices de aro**. Pero esto no pasa para cualquier escala y velocidades del sistema. Para explorar si los tamaños de cavidad, orificios y velocidades que mencionamos dan lugar a inestabilidades periódicas en el flujo (vórtices) decidimos simular un modelo crudo de laringe usando un flujo bidimensional. Las simulaciones de dinámica de fluido son computacionalmente costosas y suelen requerir el uso de *clusters* de PCs. En las últimas décadas, el método de cuadrícula de Boltzmann (método LB) surgió como un esquema competitivo para simular flujos cuasi incompresibles. El método es paralelizable y particularmente adaptable a ser programado en una unidad de procesamiento gráfico (placa de video o GPU en inglés). Una placa de video posee miles de CPUs que contrastan con los 4 u 8 que suele tener una PC convencional. Esto hace posible realizar

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

simulaciones de fluidos complejas en una PC convencional (incluso una laptop); simulaciones que tendrían tiempos de corrida prohibitivamente largos si se ejecutaran en el procesador principal. Nosotros usamos la librería Sailfish de Michal Januszewski (101) que implementa el método LB con directivas de alto nivel en Python. Usando Sailfish, armamos un modelo crudo de laringe como una caja rectangular de 1 mm de ancho y 3 mm de alto con dos aperturas de 1 mm de diámetro cada una. Usamos una cuadrícula de 150x300 con un $\Delta x = 2 \cdot 10^{-5} m/div$ un $\Delta t = 3, 3 \cdot 10^{-8} s/it$, velocidades numéricas de entre 0.01 y 0.1 (con 0.05 equivalente a 30m/s), viscosidad numérica $\nu = 1, 4 \cdot 10^{-3}$. Estos valores adimensionales de la simulación fueron elegidos de manera tal que sean compatibles con los valores físicos que ya describimos en esta sección tomando valores para la densidad y viscosidad del aire a 35° C. Describimos los resultados de estas simulaciones a continuación.

Logramos obtener oscilaciones auto sostenidas para velocidades superiores a 6 m/s. Por debajo de ese valor el flujo era uniforme y no presentaba oscilaciones. Las frecuencias obtenidas dentro de los rangos fisiológicos de velocidad abarcan el rango entre los 8 KHz y los 60 KHz compatible con el rango de las vocalizaciones ultrasónicas. La frecuencia de oscilación aumenta linealmente con la velocidad hasta que en cierto punto salta a una nueva frecuencia y continua subiendo linealmente (ver figura 6.7). Creemos que se trata de un salto entre el modo 1 y el modo 2. De ser así deberíamos ver un único vórtice dentro de la cavidad en el primer caso y dos vórtices en el segundo. Esto es efectivamente lo que observamos. En la figura 6.8 mostramos algunos cuadros que ilustran la evolución temporal del flujo en cada modo. Se observa también en la figura 6.7 un salto más, posiblemente un modo 4. Para velocidades mayores, la simulación se vuelve inestable.

6.3.1. Conclusiones

En esta hemos presentado evidencias a favor de un modelo aerodinámico para la producción de vocalizaciones ultrasónicas en la rata. Vimos que la estructura del mapa de saltos podía entenderse (al menos en parte) a partir de un modelo de silbato *bird-call* sencillo en el que el flujo se aproxima por una serie de vórtices discretos y la frecuencia del sistema está dada por un acople entre la frecuencia de emisión de vórtices y la frecuencia del sonido emitido en el segundo orificio. Para comparar con un escenario más realista, simulamos flujos en 2 dimensiones con una geometría sencilla y parámetros

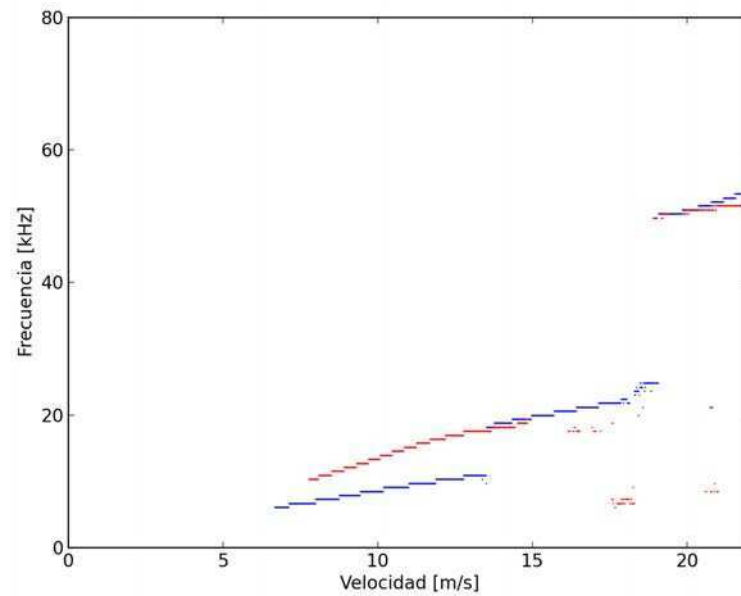


Figura 6.7: Frecuencia de oscilación en función de la velocidad - Los puntos azules corresponden a simulaciones con velocidad ascendente y los rojos a velocidad descendente. Se observa un fenómeno de histéresis pues el sistema nunca vuelve a caer al modo 1. También al ir bajando la velocidad aparece una región en la que el sistema no oscila (entre los 15 m/s y los 18 m/s).

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

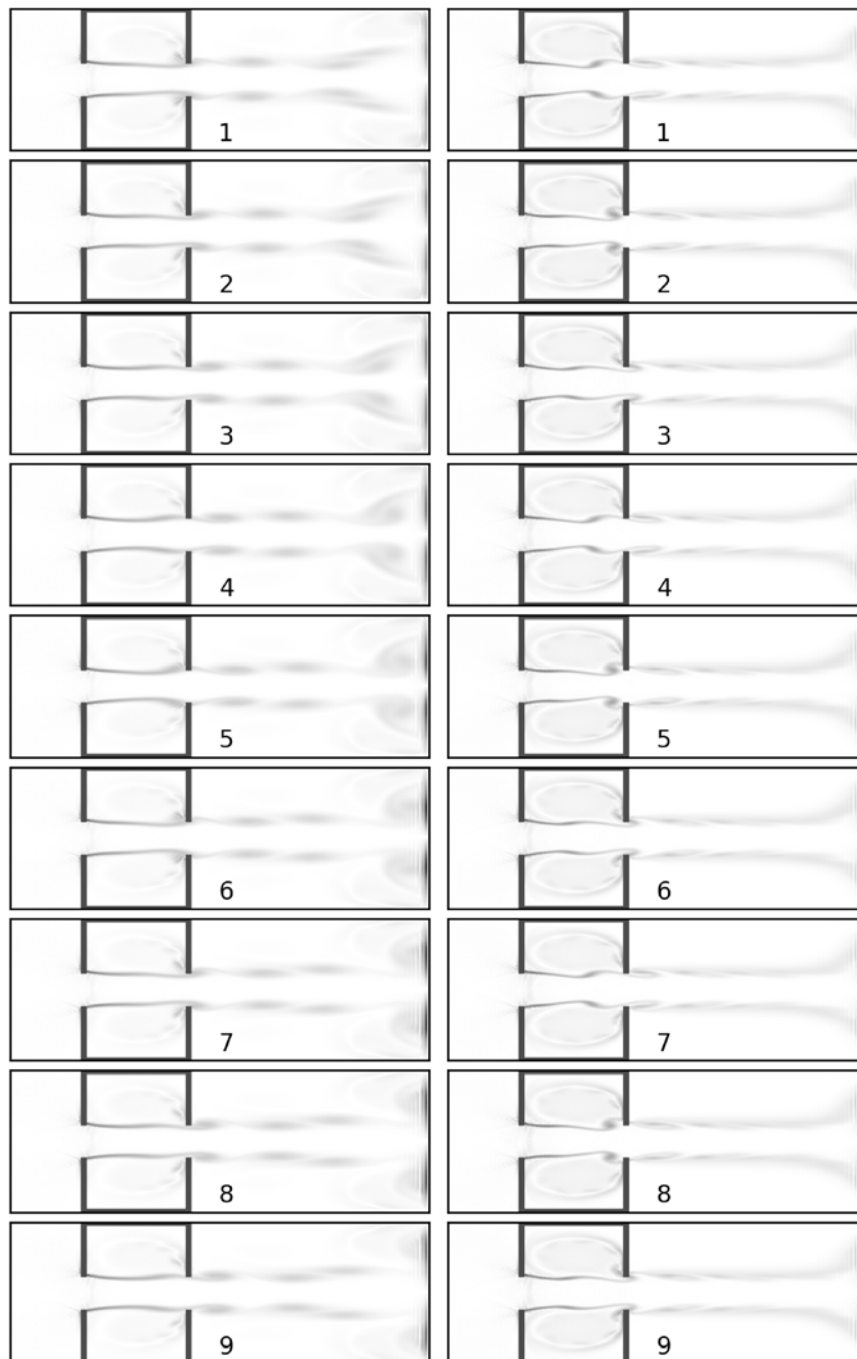


Figura 6.8: Evolución temporal del flujo - Los gráficos muestran el valor absoluto de la vorticidad a distintos tiempos. A la izquierda se observa la evolución temporal del modo 1 y a la derecha el del modo 2. Notar que a la izquierda se ve un único vórtice en cada cuadro mientras que a la derecha cuando un vórtice está impactando en el borde de la cavidad se puede ver otro cerca de la mitad

aproximados a los reales. Obtuvimos oscilaciones cuyas frecuencias están en el rango de las frecuencias típicas emitidas por las ratas y también observamos la existencia de saltos. Sería interesante construir un sistema dinámico de baja dimensionalidad que permita estudiar analíticamente las regiones de estabilidad de los distintos modos. A su vez, resultaría útil contar con datos del flujo durante la emisión de una vocalización y separar las contribuciones de U_c y de L a los cambios en frecuencia.

6.4. Control Vocal

En el capítulo introductorio de la parte de vocalizaciones (capítulo 4) estuvimos discutiendo algunos de los trabajos previos sobre el condicionamiento (clásico u operante) de vocalizaciones audibles en la rata con resultados contradictorios. Sin embargo, no hay ningún reporte de condicionamiento operante de las vocalizaciones ultrasónicas. Nosotros construimos un dispositivo experimental para entrenar a las ratas a vocalizar de manera operante frente a la presentación de un estímulo.

6.4.1. Configuración Experimental

El dispositivo consiste en una sala de entrenamiento incorporada dentro de la jaula donde se alojan los animales. A diferencia de un esquema de entrenamiento convencional, donde cada sujeto es retirado de su jaula para ser llevado a un sitio especial donde es entrenado por determinado tiempo, en nuestro dispositivo los animales pueden salir y entrar cuando quieran de la sala de entrenamiento. Esto supone una dificultad adicional ya que requiere una automatización casi total del entrenamiento pero una vez que funciona permite simplificar ampliamente la tarea de entrenamiento. A continuación sigue una descripción detallada del aparato.

La caja que contiene a todo el dispositivo y sirve para alojar a las ratas está hecha de acrílico transparente y mide 80 cm de ancho, 30 cm de alto y 50 cm de profundidad. Una separación de acrílico divide la zona de vivienda de la sala de entrenamiento que ocupa una porción menor del total de la jaula (25 cm de ancho). Dentro de la sala de entrenamiento hay una jaula de metal que se encuentra elevada del nivel del piso ya la que se accede por una pequeña escalera y orificio en el separador de acrílico. Las caras de acrílico internas a la sala de entrenamiento (que son inaccesibles a la rata pues ésta queda dentro de la jaula metálica) se encuentran insonorizadas por recubrimientos

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

de espuma acústica. También dentro de la sala de entrenamiento y exterior a la jaula metálica hay una cámara de video USB y un micrófono capacitivo ultrasónico (iguales a los usados en la arena de diálogos). Para poder detectar la presencia de una rata colocamos un par láser-fotodiodo que cruza de lado a lado y es interrumpido por el cuerpo del animal cuando está dentro de la jaula de entrenamiento. Con el objetivo de poder identificar separadamente los datos de cada individuo ubicamos un sensor de RFID (como el de la tarjeta SUBE) a la entrada de la jaula y cada animal tiene colocado un chip RFID que lo identifica. Por fuera de la jaula y en la misma dirección que la cámara y el micrófono hay un pequeño caño metálico (de 0,5 mm de diámetro) que viene del exterior de toda la jaula acrílica y permite entregar una recompensa (ya sea agua o agua azucarada). Del lado externo, el tubo es de plástico y viene de un recipiente elevado que funciona como tanque. El recorrido del tubo plástico se encuentra interrumpido por una electro-válvula que permite controlar la dosis de la recompensa. Nuevamente dentro de la sala, formando una cruz entre el final del caño metálico y la jaula hay dos pares de láser-fotodiodo que son interrumpidos por la rata al buscar la recompensa. A la altura del micrófono, hay distintos leds de colores que servirán como claves visuales para del comportamiento operante. Las señales de cámara y micrófono se conectan directamente a la PC, al puerto USB y a una placa de adquisición A/D respectivamente. Todas las demás señales de: sensor RFID, fotodiodos, leds y electro-válvula son controladas por una placa Arduino que se encuentra adosada a la cara externa de toda la jaula acrílica. La placa Arduino se conecta a la PC mediante el puerto USB. Un esquema de todo esto puede verse en la figura 6.9.

Hay dos elementos de software que controlan el desarrollo del experimento. Código de MATLAB controla desde la PC el funcionamiento general del experimento y almacena los datos mientras que código de C en la placa Arduino controla el desarrollo de cada ensayo (trial). Ambos se comunican serialmente en el intervalo entre trials. A continuación describimos el flujo de funcionamiento. Para simplificar la descripción en lugar de decir: el código dentro de la placa Arduino diremos simplemente Arduino y en lugar de: el código de MATLAB diremos sencillamente MATLAB. Empezamos la descripción con la sala de entrenamiento vacía. Arduino chequea periódicamente el estado del haz de láser que detecta la presencia de una rata. Si una rata ingresa a la sala, pide al sensor RFID el código que identifica al individuo y se lo pasa vía comunicación serial a MATLAB. MATLAB selecciona el programa de entrenamiento adecuado para

ese individuo y le pasa a Arduino los parámetros del próximo trial. Más adelante, cuando describamos el protocolo, detallaremos cuales son esos parámetros. Arduino ejecuta el trial y, según la respuesta de la rata, entrega o no una recompensa. Dado que las respuestas son vocalizaciones, quien monitorea permanentemente y en tiempo real para decidir si hubo o no una vocalización es MATLAB y envía una señal por un pin digital a Arduino cuando esto ocurre (esto es más rápido que la comunicación por puerto serial).

El desarrollo de un trial es el siguiente. Se enciende una de dos luces posibles de acuerdo al tipo de trial (S_D o S_Δ) que permanecerá encendida por un máximo de 5 segundos. Si se trata de un trial S_D la rata debe vocalizar. Si esto ocurre, se apaga la luz, se escuchan 3 clics y la válvula entrega una recompensa (una gota de agua o agua con azúcar). Si la rata no responde dentro de los 5 segundos, la luz se apaga y no hay recompensa. En el caso de un trial tipo S_Δ , se prende otra luz y la contingencia es al revés. Si la rata vocaliza se apaga la luz y no da recompensa y si la rata permanece en silencio los 5 segundos recibe la recompensa al final. Los trials están separados por una pausa de 250 ms. Para fomentar que las ratas ingresen en la cámara de entrenamiento se ofrecen también recompensas libres en forma aleatoria entre trial y trial. Las recompensas son menores (en volumen de agua) a las que corresponden a responder un trial correctamente y están acompañadas por un único clic. La probabilidad de recompensas libres se corrige automáticamente de manera tal de asegurar que la rata recibe un piso mínimo de recompensas por minuto y se hace cero en cuanto el desempeño de la rata mejora y obtiene sus recompensas como resultado de haber respondido correctamente. La probabilidad de que se presente un trial S_D o un S_Δ depende de las respuestas de la rata; de manera tal, que recibe una proporción mayor del tipo de trials en el que haya cometido más errores. Esto garantiza que no es una estrategia válida responder siempre igual y recibir la mitad de las recompensas.

Para verificar que nuestro dispositivo funciona decidimos entrenar antes que el comportamiento vocal un comportamiento motor que sabemos puede ser condicionado de manera operante: el *nose poking* (que consiste en que la rata introduce su nariz en un orificio). Para esto, agregamos a la jaula una cavidad con un led infrarrojo y un fotodiodo (lo que se conoce como un *nose port*). Todo lo que describimos anteriormente permanece igual, la única diferencia es que la respuesta es motora en lugar de vocal.

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

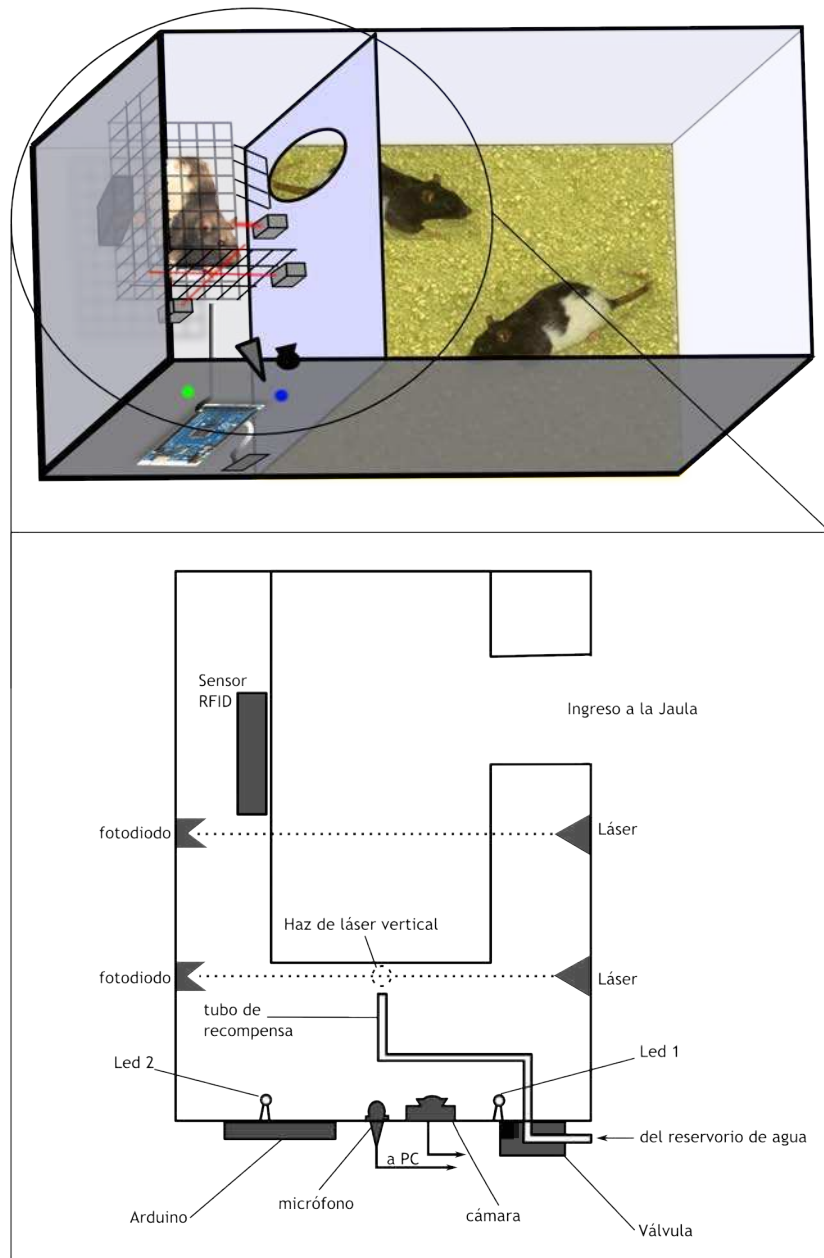


Figura 6.9: Configuración del dispositivo experimental - No se indican las conexiones desde y hacia la placa Arduino para no afectar la legibilidad del dibujo. Todos los elementos excepto el micrófono y la cámara están conectados a la placa. Ver texto para una descripción detallada.

6.4.2. Resultados

En primera instancia presentamos los resultados del entrenamiento de una acción motora que sabemos que puede ser condicionada operantemente: el *nose poking*. Entrenamos con este protocolo a dos individuos. Resumimos, en la figura 6.10, el desempeño de uno de ellos en las distintas etapas del entrenamiento y a medida que la tarea se va haciendo cada vez más difícil. La figura está dividida en distintas etapas. En la etapa 1, sólo se presentan trials del tipo S_D y se puede ver como la rata aumenta su proporción de respuestas correctas (y por lo tanto disminuye la cantidad de trials con recompensa gratis). En la etapa 2, se introducen los trials S_Δ y se observa una caída inicial del desempeño seguido de una paulatina mejora. En la etapa 3, se acorta el tiempo máximo de duración del trial de 10 s a 5 s. Nuevamente, se observan caídas en el desempeño hasta que finalmente la fracción de correctos se estabiliza (en la etapa 4) alrededor del 70 %, aunque con variaciones considerables. Mostramos, también en la misma figura, histogramas de tiempo de respuesta para los trials S_D (los trials en los que la rata debe responder). Ahí se puede ver como en los primeros trials la distribución es aproximadamente exponencial y no hay un tiempo de respuesta claro mientras que en los últimos se observa un pico definido. Es llamativa la variabilidad del desempeño en la etapa final, una vez que ya aprendieron la tarea. Es probable que ésta se deba a que, como las ratas pueden entrar y salir a voluntad de la sala de entrenamiento, no siempre que entren en ella será en busca de recompensa. Sin embargo, el haz de láser detecta, de todas formas, la presencia de una rata y comienza a presentar trials a los cuales la rata puede no estar prestando atención. Para solucionar esto, decidimos usar los haces láser que se encuentran frente al tubo de recompensas (que sirven para saber si la rata buscó la recompensa o no) como forma de que la rata inicie los trials. Es decir que la rata tiene que interrumpir los haces del tubo de recompensa por un cierto tiempo para iniciar un trial. El aprendizaje de este nuevo esquema se muestra en la figura 6.11 (hasta el trial 10000).

Hasta ahora todo parece indicar que el protocolo de entrenamiento funciona correctamente. Una última prueba que se suele hacer para mostrar el comportamiento operante es invertir la regla. Esto se muestra también en la figura 6.11, donde en el trial 10000 se procede a dar vuelta la contingencia. A partir de ese trial, el estímulo asociado a S_D pasa a ser el de S_Δ y viceversa. Se puede ver como el desempeño cae rápidamente

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

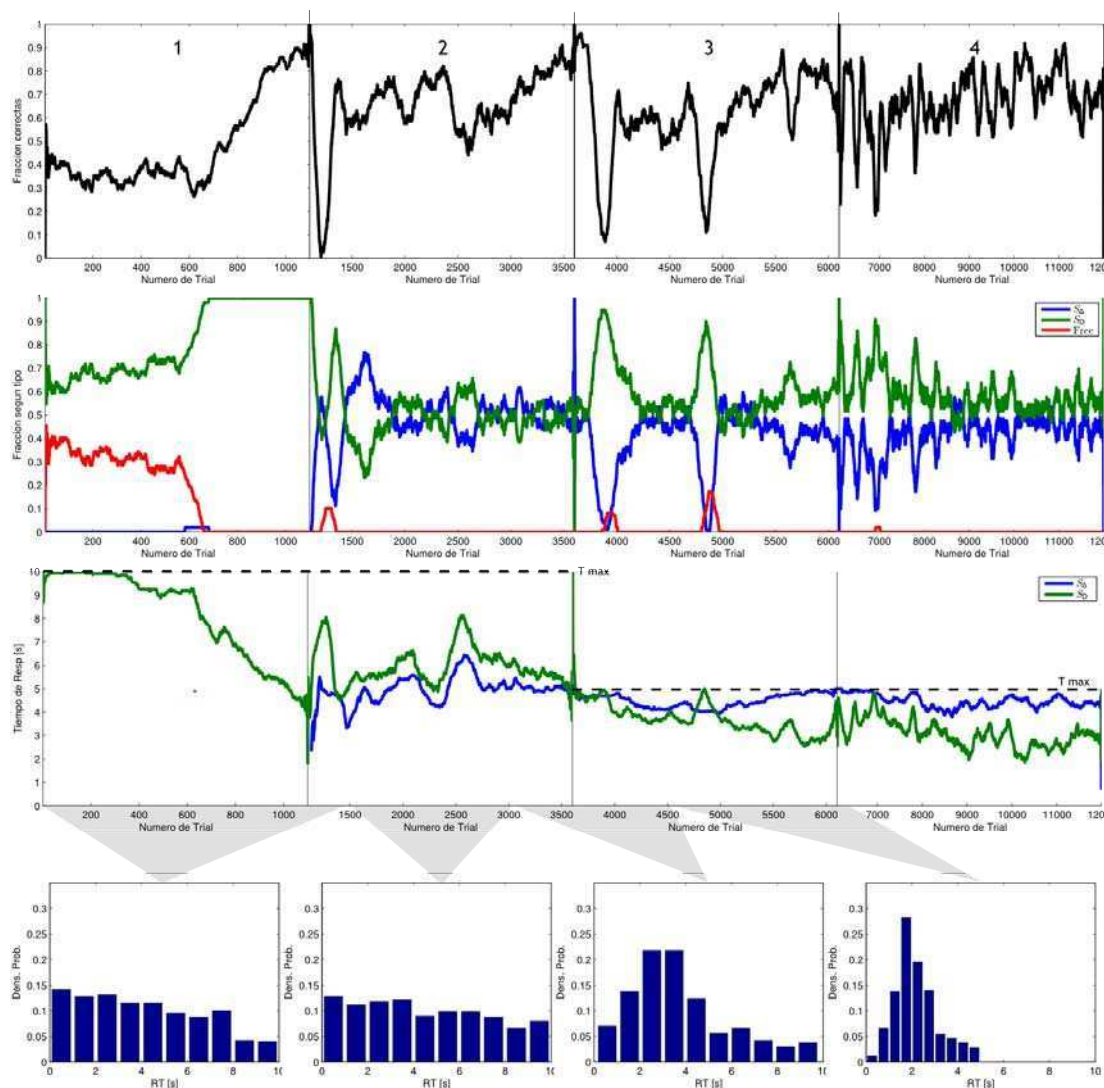


Figura 6.10: Entrenamiento de una Rata - La evolución del entrenamiento se muestra de manera continua pero dividida en distintas escalas temporales para apreciar diferentes fases de la misma. Las tres primeras filas corresponden a tres medidas del desempeño de la rata a medida que avanzan los trials. En la primera fila, mostramos la fracción de trials correctos; en la segunda, la fracción de trials presentados separados según su tipo y, finalmente, en la tercera fila, presentamos los tiempos de respuesta separados también según el tipo de trial. En la última de las filas mostramos la distribución de tiempos de respuesta para trials del tipo S_D en distintos momentos del entrenamiento (indicados con sombra). En la etapa 1, sólo se presentan trials del tipo S_D y se puede ver como la rata aumenta su proporción de respuestas correctas (y por lo tanto disminuye la cantidad de trials con recompensa gratis). En la etapa 2, se introducen los trials S_Δ y se observa una caída inicial del desempeño seguido de una paulatina mejora. En la etapa 3, se acorta el tiempo máximo de duración del trial de 10 s a 5 s. Nuevamente, se observan caídas en el desempeño hasta que finalmente la fracción de correctos se estabiliza (en la etapa 4) alrededor del 70% aunque con variaciones considerables.

para luego recuperarse paulatinamente a niveles previos a la manipulación.

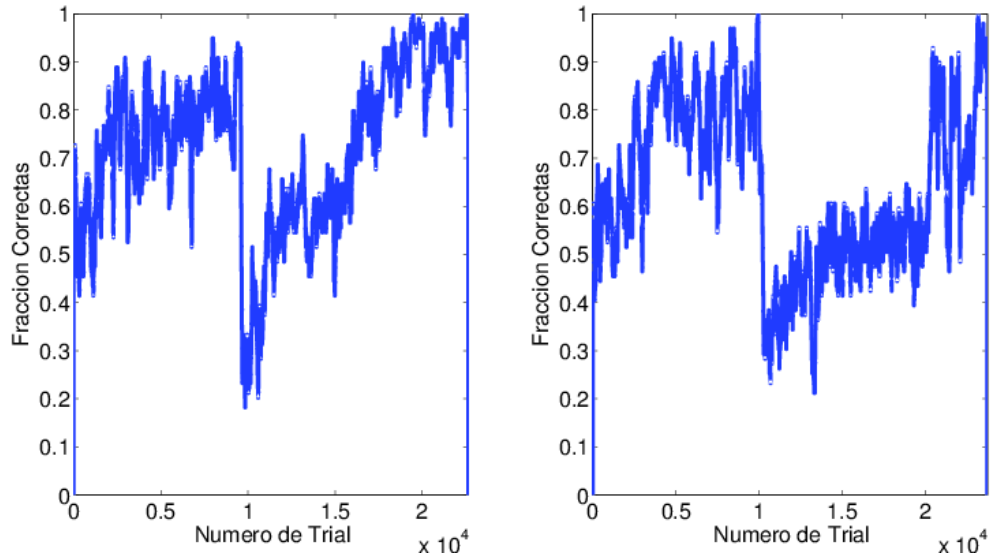


Figura 6.11: Inversión de la regla - Desempeño para los dos animales entrenados en la modalidad de inicio voluntario de trials. Se muestra también lo que ocurre cuando, en el trial 10000, se invierte la regla; es decir, el estímulo asociado a S_D pasa a ser el de S_Δ y viceversa.

Ahora sí, estamos en condiciones de observar qué ocurre si repetimos este protocolo de entrenamiento no ya para una acción motora sino para una vocalización. En la figura 6.12 presentamos la proporción de respuestas correctas para la primera fase de entrenamiento, es decir, cuando solamente se presentan trials del tipo S_D . En esta fase esperaríamos que la tasa de vocalizaciones aumente a medida pasan los trials, lo cual se vería reflejado en un aumento de la proporción de aciertos. Sin embargo, se puede ver que, a diferencia de lo que ocurre en la primera fase de la figura 6.10, las ratas no desarrollan un comportamiento consistente. Muchos de los individuos presentan picos con momentos de alta tasa de vocalización. Esto se debe a que, dado que todas las ratas viven juntas en la jaula anexa a la sala de entrenamiento, cuando se producen momentos de interacción social fuerte (juego o pelea) esto estimula la producción vocal de la rata que se encuentra en ese momento en la sala de entrenamiento. En principio, ésta es una característica deseable ya que sirve para fomentar el comportamiento que queremos llevar bajo control operante. Varios individuos presentan altas tasas de vocalización al comienzo, cuando recién fueron puestas en la jaula y hay una interacción social más intensa entre ellos. Los únicos dos sujetos que presentaron un comportamiento

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

relativamente consistente y tasas de vocalización sostenidas en el tiempo son R4 y R6. Sin embargo, sus comportamientos distan de lo observado en el caso de la acción motora donde alrededor del trial número 1000 ya se obtenían respuestas correctas en el 90 % de los trials.

Por estas razones, solo R4 y R6 fueron expuestos a trials del tipo S_{Δ} . Los resultados de esa etapa posterior se muestran en la figura 6.13. El desempeño de ambas ratas se mantiene estable y en valores bajos. Nada parece indicar que haya signos de aprendizaje. No hay diferencias significativas en los tiempos de respuesta para trials del tipo S_D (aciertos) y para trials S_{Δ} (errores) excepto al comienzo y solo en el caso de R4. Esto se debe a que a R4 (la primera rata en ser presentada con trials S_{Δ}) se le presentaron inicialmente trials S_{Δ} que no estaban atados a la contingencia; es decir, que no eran recompensados por no vocalizar durante ellos. Esto se hizo porque, dado que las ratas están, naturalmente, relativamente calladas, al introducir trials S_{Δ} comenzarían a recibir recompensas que podrían ser percibidas como “gratis”.

A pesar de que no logran producir el comportamiento de manera operante, es llamativo que R4 y R6 lograron de alguna forma conservar una tasa de vocalizaciones razonable como para obtener algunas recompensas. Más aún, el tiempo de respuesta, si bien no distingue entre ambas condiciones, es relativamente estable; está, de alguna forma, atado al inicio del trial. ¿Cómo es que logran esto sin tener un control voluntario? Veremos que la vocalización parece estar atada a una respuesta emocional provocada por la recompensa. Para estudiar esto, podemos hacer un histograma de los tiempos de respuesta para 4 condiciones distintas: si los trials son S_D o S_{Δ} y si el trial anterior fue recompensado o no. Esto es lo que mostramos en la figura 6.14. Ahí se puede ver como la distribución presenta un pico solo en los casos en los que el trial fue precedido por uno recompensado, indistintamente de si se trata de un S_D o un S_{Δ} . En efecto, esto es consistente con la interpretación de la vocalización como una respuesta afectiva positiva frente a la experiencia placentera de la recompensa.

Toda la evidencia presentada hasta ahora indica que las vocalizaciones son, en efecto, muy distintas de otros gestos motores (como por ejemplo el *nose poking*) y que no pueden ser puestas bajo control operante. Cinco de los siete animales entrenados simplemente no pudieron incrementar su tasa de vocalizaciones para conseguir una recompensa a pesar de haber sido expuestos a decenas de miles de contingencias incluso en situaciones sociales, dadas por la cohabitación con otras ratas, que los llevaban

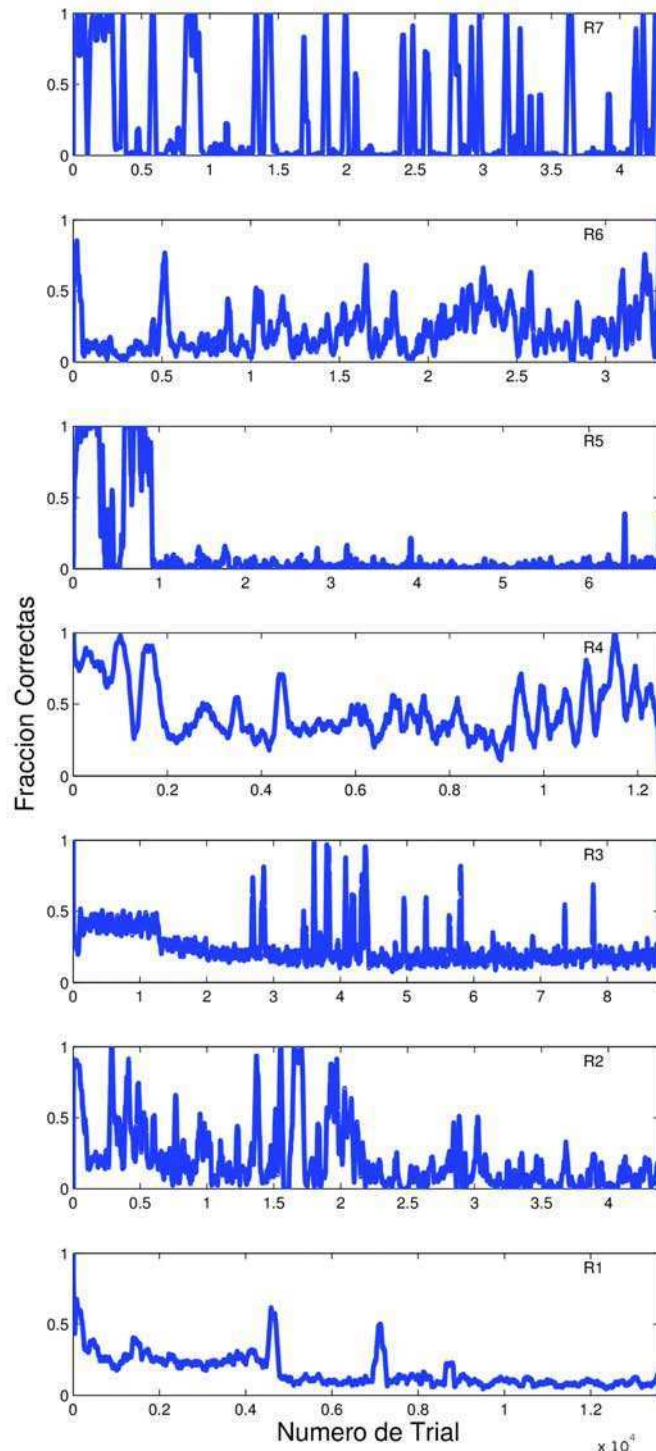


Figura 6.12: Entrenamiento Vocal - Desempeño para siete ratas en la primera fase de entrenamiento. Todos los trials son S_D , es decir que los sujetos deben vocalizar lo más posible para maximizar las recompensas. Sin embargo, no se observa un comportamiento de vocalización consistente. Solo R4 y R6 mostraron una tasa de vocalizaciones relativamente sostenida en el tiempo y fueron, entonces, presentados con trials S_Δ

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

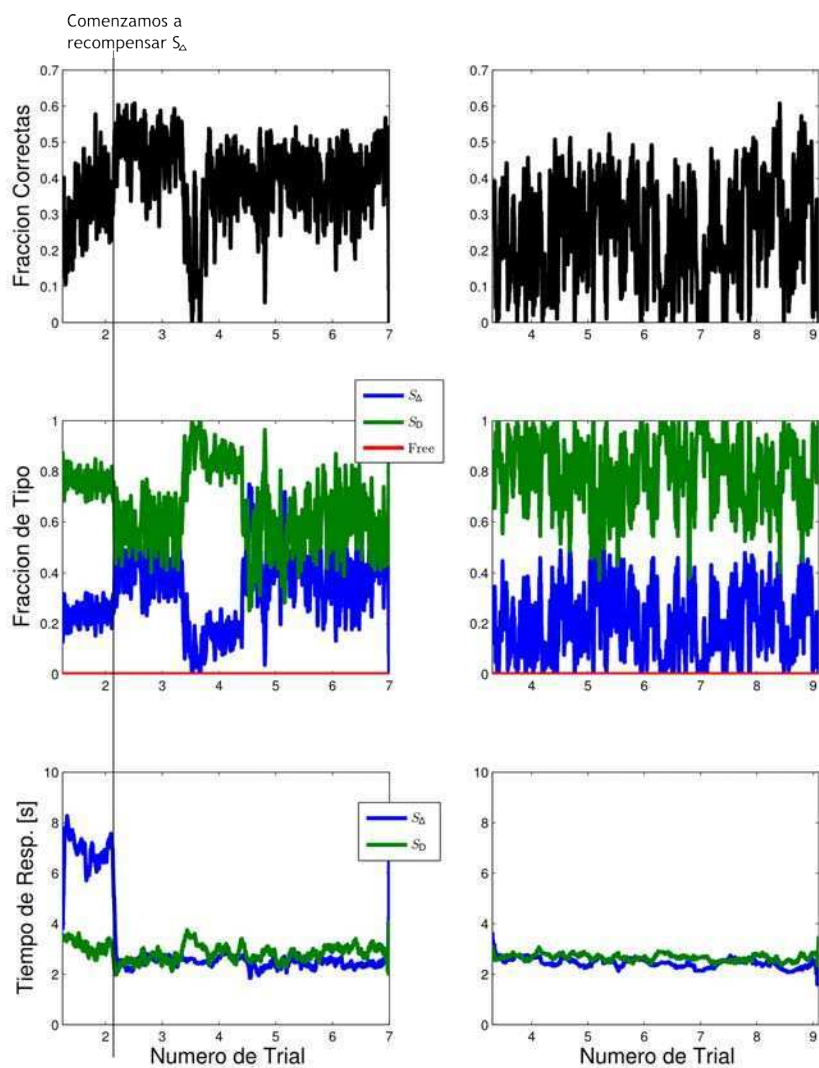


Figura 6.13: Entrenamiento Vocal segunda fase - R4 (paneles de la izquierda) y R6 (derecha) son expuestos a Trials del tipo S_{Δ} . R4 fue el primer sujeto en ser expuesto a este tipo de trials y en primera instancia los presentamos sin contingencia, es decir, no recompensaban por ser completados correctamente. Luego, según se indica en la figura, comenzamos a recompensar los trials S_{Δ} . R6 fue directamente expuesto a los trials en conjunto con la contingencia.

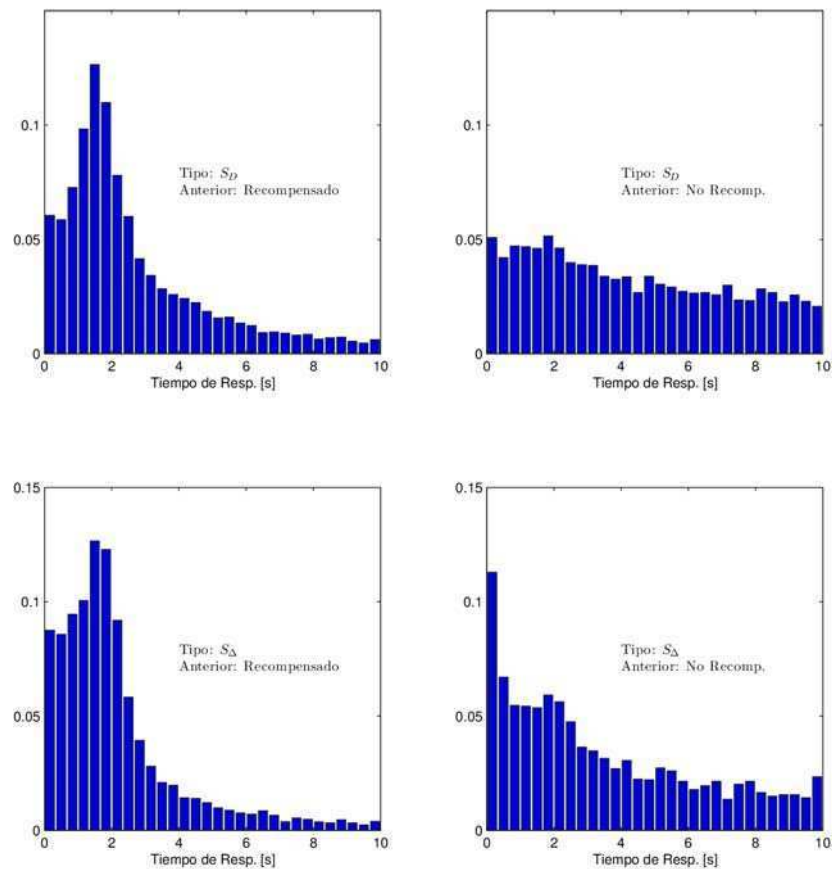


Figura 6.14: Histogramas de tiempo de Respuesta según distintas condiciones - Todos los histogramas corresponden a R4. Los tiempos de respuesta están separados de acuerdo al tipo de trial en el que fueron medidos. Se los separa también por el desempeño de la rata en el trial anterior; es decir, si recibió o no recibió recompensa. Se puede observar que los trials posteriores a recibir una recompensa (indistintamente de que tipo de trial se trate) presentan un tiempo de respuesta atado al inicio del trial. En cambio, esto no ocurre si el trial anterior no fue recompensado.

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

a vocalizar naturalmente y recibir recompensas. Los dos individuos que si lograron vocalizar establemente en proporciones cercanas a la mitad de los trials no pudieron ajustar ese comportamiento a los cambios en un estímulo externo; es decir, atar sus vocalizaciones a los momentos S_D y no a los S_Δ . Más aún, en estos individuos vimos que la respuesta de vocalización ocurría un cierto tiempo después de recibir una recompensa independientemente del tipo de trial que fuera presentado.

Algo que nos llamó la atención, sin embargo, es que el tipo de vocalizaciones que hacían, tanto R4 como R6, hacia el final del experimento eran muy estereotipadas. En la figura 6.15 mostramos ejemplos de vocalizaciones de R6 al comienzo del experimento y vocalizaciones del final. Se puede ver mucha más variabilidad en las primeras, mientras que la segundas son vocalizaciones con poca modulación y de frecuencia similar. Esto no se debe sin embargo a una deriva “natural” del tipo de vocalizaciones con el tiempo, pues, terminado el experimento, grabamos vocalizaciones en la arena de dialogo y se vuelve a obtener mayor riqueza (también mostrado en la figura 6.15). El mismo efecto se puede ver como un cambio gradual en la composición de las nubes en los mapas de propiedades como los que usamos en el capítulo pasado (ver figura 6.16).

No sabemos a que se debe esta deriva pero es posible que refleje un intento de alterar las variables que sí pueden ser controladas voluntariamente para incrementar la probabilidad de una vocalización. Algunas de estas variables podrían ser: la presión de los pulmones, la forma del tracto vocal, la postura del torso, la de la cabeza. Esto podría favorecer la producción de vocalizaciones que sean artificiales, y con propiedades acústicas distintas a las vocalizaciones “naturales”.

6.4.3. Conclusiones

En esta sección exploramos el grado de control voluntario de las ratas durante la emisión de vocalizaciones ultrasónicas. Para esto construimos un dispositivo de entrenamiento de 24 hs con dos modos de funcionamiento: uno motor y otro vocal. Usando este dispositivo nos fue posible entrenar fácilmente un gesto motor (*nose poking*) de manera operante. Las ratas ejecutaban el gesto frente a la presentación de un estímulo y lo evitaban frente a la presentación de otro con tasas de éxito cercanas al 90%. Más aún, nos fue posible invertir la regla y, luego de cierto tiempo, recuperar valores de desempeño similares a los que había previamente a esta manipulación. A pesar del éxito del dispositivo para entrenar un gesto motor, no nos fue posible usar el mismo

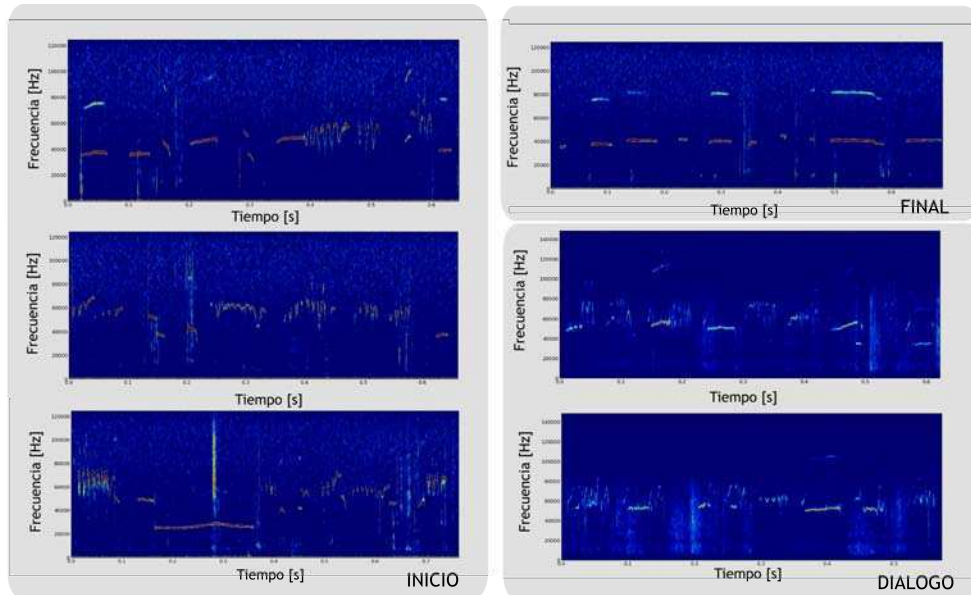


Figura 6.15: Ejemplos de Vocalizaciones - Ejemplos de vocalizaciones de R6 en distintos momentos. INICIO: vocalizaciones al inicio del experimento. FINAL: Vocalizaciones del final del experimento. DIALOGO: vocalizaciones luego de terminado el experimento grabadas en el arena de diálogos.

protocolo para entrenar vocalizaciones de manera operante. De las siete ratas entrenadas solo dos pudieron mantener un nivel estable de vocalizaciones pero ninguna de las dos pudo emitir las de manera contingente a un estímulo externo. Todo esto parece indicar que el grado de control voluntario de las vocalizaciones ultrasónicas es muy bajo o inexistente.

6. OTROS ASPECTOS DE LA PRODUCCIÓN VOCAL EN RATAS

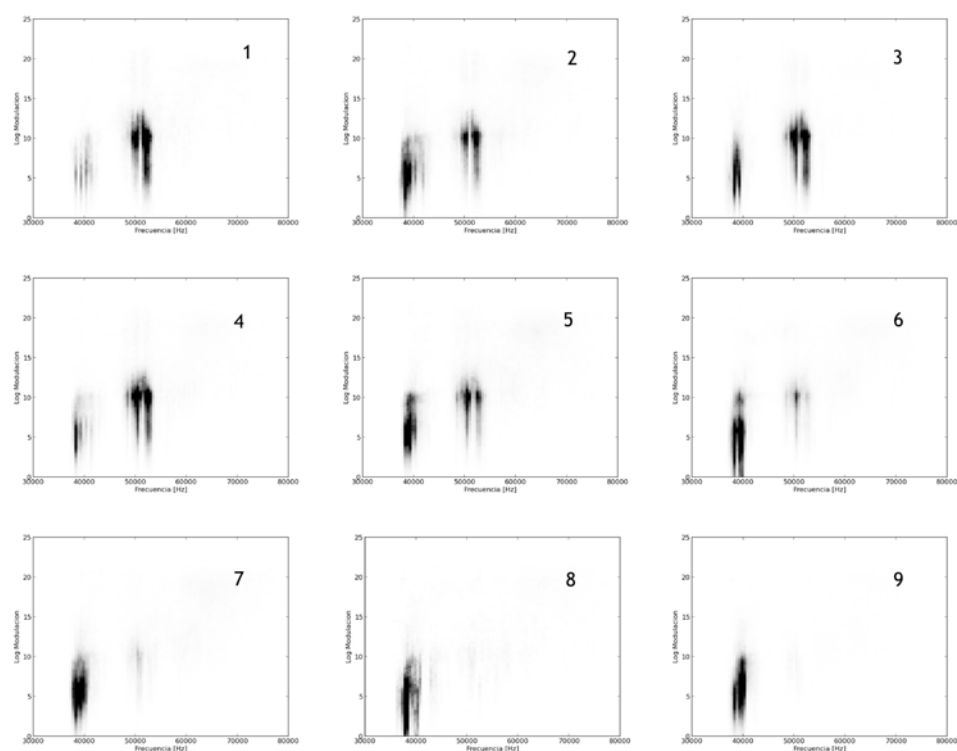


Figura 6.16: Cambios en los mapas de R6 - Cambios en los mapas de R6 a lo largo del entrenamiento. Toda la secuencia de trials completa mostrada en las figuras 6.12 y 6.13 se presenta aquí dividida en 9 intervalos temporales de igual duración. Se puede ver como cambia la conformación de las nubes.

7

Cierre

Hemos llegado al final del viaje. Dimos una vuelta al bucle de la comunicación recorriendo cada una de las tres grandes áreas Jamesianas y aprendimos algunas cosas en el camino. Empezamos estudiando una puerta de entrada: el reconocimiento de la palabra hablada. Analizamos una ilusión que se produce frente a la escucha de una palabra repetida y pudimos mostrar que cumple con todas las características de un fenómeno de percepción biestable. Esto pone de manifiesto cabalmente cómo la estrategia de percibir estímulos ambiguos alternando las hipótesis más probables, es una estrategia muy general de nuestros cerebros. A la vez, pensar el reconocimiento de palabras como fenómeno biestable nos permitió sugerir que hay un elemento faltante en los modelos tradicionales: la adaptación. La adaptación es la que permite que el percepto ganador de la competencia, disminuya su actividad lentamente para dejar paso al percepto que le siga en nivel de plausibilidad.

Siguiendo nuestro recorrido, pasamos al procesamiento y elaboración del estímulo (pensamiento). Allí estudiamos algunos aspectos de la forma de organización semántica de los conceptos en nuestras mentes. Usamos el texto escrito como ancla para definir algorítmicamente una noción de significado a partir de la estadística de co-ocurrencias entre palabras. Esto nos permitió estudiar de manera cuantitativa la evocación espontánea de un pensamiento por otro (asociación libre). Tratamos de entender este proceso como una caminata aleatoria en un espacio definido por la estadística de textos y vimos que ningún modelo de vecinos acotado es capaz de capturar esa dinámica. Tu- vimos que recurrir entonces a modelos libres de escala. Ocurre que al asociar libremente nos permitimos hacer algunas conexiones de largo alcance que no son comunes en el

7. CIERRE

texto escrito.

Finalmente, llegamos a la última etapa que corresponde al agente efectuando un cambio en el ambiente; en nuestro caso, eso corresponderá a vocalizar. Hicimos aquí un gran salto y pasamos a estudiar las vocalizaciones no en humanos sino en ratas. Definimos propiedades acústicas tiempo a tiempo que nos permitieron reconocer segmentos estereotípicos que componen cada vocalización. Luego, vimos que las transiciones entre esos segmentos no eran aleatorias sino que había secuencias que se repetían más de lo que esperaríamos por azar. Esto nos permitió recuperar algunas de las categorías definidas ad-hoc por otros grupos de investigación pero en forma justificada por la estructura emergente de los datos. Armados con nuestras categorías pudimos estudiar cómo éstas se estructuran en el tiempo y vimos que hay un efecto de orden y que el sistema tiene memoria. Todo esto sugiere que si bien no parece haber reglas estrictas hay una estructuración probabilística de las categorías en el tiempo. Pudimos observar, además, interacciones entre individuos que se hacen visibles en una correlación negativa de sus tasas de vocalización a escalas temporales cortas.

Adicionalmente, pudimos explorar otros aspectos de las vocalizaciones ultrasónicas. Vimos que hay una fuerte correlación entre la locomoción y la tasa de vocalización pero, también, que no es la vocalización un simple subproducto de la caminata. Pudimos explorar, también, los mecanismos físicos de producción sonora. Observamos que los saltos de frecuencia presentes en las vocalizaciones corresponden a modos de funcionamiento de una cavidad resonante y mostramos que eso es consistente con modelos de silbato aerodinámico para la producción. Logramos simular un modelo aerodinámico sencillo con magnitudes realistas y compatibles con la anatomía de la rata que produjo perturbaciones periódicas en el rango de frecuencias esperado.

En un último experimento, evaluamos el grado de control voluntario que las ratas poseen durante la emisión de vocalizaciones ultrasónicas. Para esto diseñamos una jaula de entrenamiento de 24 hs totalmente automatizada. El protocolo establecido nos permitió entrenar operantemente un gesto motor (*nose poking*) pero no nos fue posible hacer lo mismo con las vocalizaciones. Todo parece indicar que el grado de control voluntario de las vocalizaciones es limitado.

Bibliografía

- [1] R.M. Warren and R.L. Gregory. An auditory analogue of the visual reversible figure. *The American Journal of Psychology; The American Journal of Psychology*, 1958. 7, 8, 9
- [2] R.M. Warren. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *British Journal of Psychology*, 52(3):249–258, 1961. 7, 9
- [3] L.A. Necker. Lxi. observations on some remarkable optical phænomena seen in switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid. 1832. 7
- [4] Philipp Sterzer, Andreas Kleinschmidt, and Geraint Rees. The neural bases of multistable perception. *Trends in cognitive sciences*, 13(7):310–318, 2009. 7
- [5] Mingzhou Ding, Betty Tuller, and JA Kelso. Characterizing the dynamics of auditory perception. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):70–75, 1995. 9
- [6] Tuller Ditzinger, B Tuller, and JAS Kelso. Temporal patterning in an auditory illusion: the verbal transformation effect. *Biological cybernetics*, 77(1):23–30, 1997. 9
- [7] James L McClelland and Jeffrey L Elman. The trace model of speech perception. *Cognitive psychology*, 18(1):1–86, 1986. 9
- [8] Dennis Norris. Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3):189–234, 1994. 9

BIBLIOGRAFÍA

- [9] Dennis Norris, James M McQueen, and Anne Cutler. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23(3):299–324, 2000. 9
- [10] Paul A Luce and David B Pisoni. Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1):1, 1998. 9
- [11] Paul A Luce, Stephen D Goldinger, Edward T Auer, and Michael S Vitevitch. Phonetic priming, neighborhood activation, and parsyn. *Perception & psychophysics*, 62(3):615–625, 2000. 9
- [12] James A Bashford Jr, Richard M Warren, and Peter W Lenz. Polling the effective neighborhoods of spoken words with the verbal transformation effect. *The Journal of the Acoustical Society of America*, 119(4):EL55, 2006. 9
- [13] Marc Sato, Monica Baciú, H el ene L evenbruck, Jean-Luc Schwartz, Marie-Agnes Cathiard, Christoph Segebarth, and Christian Abry. Multistable representation of speech forms: A functional mri study of verbal transformations. *NeuroImage*, 23(3):1143–1151, 2004. 9
- [14] Hirohito M Kondo and Makio Kashino. Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage*, 36(1):123–130, 2007. 9
- [15] Anahita Basirat, Marc Sato, Jean-Luc Schwartz, Philippe Kahane, and Jean-Philippe Lachaux. Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *Neuroimage*, 42(1):404–413, 2008. 10
- [16] David Alais and Randolph Blake. *Binocular rivalry*. MIT Press, 2004. 10
- [17] Nicholas J Wade. *A natural history of vision*. MIT Press, 2000. 11
- [18] Willem JM Levelt. *On binocular rivalry*, volume 2. Mouton, 1968. 11
- [19] Jean-Michel Hup e and Nava Rubin. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision research*, 43(5):531–548, 2003. 11, 13
- [20] David A Leopold and Nikos K Logothetis. Multistable phenomena: changing views in perception. *Trends in cognitive sciences*, 3(7):254–264, 1999. 12

-
- [21] Sidney R Lehky. Binocular rivalry is not chaotic. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259(1354):71–76, 1995. 13
- [22] YH Zhou, JB Gao, KD White, I Merk, and K Yao. Perceptual dominance time distributions in multistable visual perception. *Biological cybernetics*, 90(4):256–263, 2004. 13
- [23] Q-q plot. *Wikipedia*, <http://en.wikipedia.org/wiki/Q17>
- [24] Daniel Pressnitzer and Jean-Michel Hupé. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13):1351–1357, 2006. 18
- [25] Ming Meng and Frank Tong. Can attention selectively bias bistable perception? differences between binocular rivalry and ambiguous figures. *Journal of vision*, 4(7):2, 2004. 21
- [26] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum. *The measurement of meaning*. Univ of Illinois Pr, 1957. 23, 24
- [27] K. Lund and C. Burgess. Producing High-Dimensional Semantic Spaces From Lexical Co-Occurrence. *Behavior Research Methods Instruments and Computers*, 28:203–208, 1996. 24
- [28] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990. 24
- [29] D.P. Spence and K.C. Owens. Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19(5):317–330, 1990. 24
- [30] W.A. Bousfield. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49:229–240, 1953. 24
- [31] J.J. Jenkins, W.D. Mink, and W.A. Russell. Associative clustering as a function of verbal association strength. *Psychological Reports*, 4:127–136, 1958. 24
- [32] J. Deese. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1):17–22, 1959. 24

BIBLIOGRAFÍA

- [33] J. Deese. *The Structure of Associations in Language and Thought*. Johns Hopkins University Press, 1965. 24
- [34] D.L. Nelson, T.A. Schreiber, and C.L. McEvoy. Processing implicit and explicit representations. *Psychological Review*, 99(2):322–348, 1992. 24
- [35] D.L. Nelson, N. Zhang, and V.M. McKinney. The ties that bind what is known to the recognition of what is new. *Learning, Memory*, 27(5):1147–1159, 2001. 24, 25
- [36] M. Steyvers, R.M. Shiffrin, and D.L. Nelson. Word association spaces for predicting semantic similarity effects in episodic memory. *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. Washington, DC: American Psychological Association, 2004. 24
- [37] A.M. Collins and E.F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975. 25
- [38] J.L. McClelland and E. Jenkins. Nature, nurture, and connections: Implications of connectionist models for cognitive development. *Architectures for intelligence*, pages 41–73, 1991. 25
- [39] C. Burgess and K. Lund. Multiple constraints in syntactic ambiguity resolution: A connectionist account of psycholinguistic data. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pages 90–95, 1994. 25
- [40] R.F. Cancho. The small world of human language. *Proceedings of the Royal Society B: Biological Sciences*, 268(1482):2261–2265, 2001. 25
- [41] M. Sigman and G.A. Cecchi. Global organization of the Wordnet lexicon. *Proceedings of the National Academy of Sciences*, 99(3):1742–1747, 2002. 25
- [42] R.F. Cancho and R.V. Sole. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791, 2003. 25
- [43] M. Steyvers and J.B. Tenenbaum. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29(1):41–78, 2005. 25

-
- [44] E.W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959. 25
- [45] DL Nelson and N. Zhang. The ties that bind what is known to the recall of what is new. *Psychon Bull Rev*, 7(4):604–17, 2000. 25
- [46] S. Hotta, K. Inoue, and K. Urahama. Extraction of Fuzzy Clusters from Weighted Graphs. *Electronics and Communications in Japan(Part III Fundamental Electronic Science)*, 86(1):80–88, 2003. 28
- [47] J.W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409, 1969. 29
- [48] Benoit B Mandelbrot. *The fractal geometry of nature*. Macmillan, 1983. 35
- [49] John W Anderson. The production of ultrasonic sounds by laboratory rats and other mammals. *Science*, 119(3101):808–809, 1954. 42
- [50] LH Roberts. The rodent ultrasound production mechanism. *Ultrasonics*, 13(2):83–88, 1975. 43
- [51] Laurence H Roberts. Evidence for the laryngeal source of ultrasonic and audible cries of rodents. *Journal of Zoology*, 175(2):243–257, 1975. 43
- [52] Laurence H Roberts. Correlation of respiration and ultrasound production in rodents and bats. *Journal of Zoology*, 168(4):439–449, 1972. 43
- [53] LAURENCE H ROBERTS. The functional anatomy of the rodent larynx in relation to audible and ultrasonic cry production. *Zoological Journal of the Linnean Society*, 56(3):255–264, 1975. 43
- [54] Weisz D.J. Yang B.Y. Fung K. Amirali A. Sanders, I. The mechanism of ultrasonic vocalization in the rat. *Soc. Neurosci. Abstr.* 27, page 88.19, 2001. 43, 91
- [55] Christine V Portfors. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science*, 46(1):28–34, 2007. 43

BIBLIOGRAFÍA

- [56] Jennifer M Wright, Jim C Gourdon, and Paul BS Clarke. Identification of multiple call categories within the rich repertoire of adult rat 50-khz ultrasonic vocalizations: effects of amphetamine and social context. *Psychopharmacology*, 211(1):1–13, 2010. 44, 45
- [57] Samuel Anthony Barnett. *The rat: A study in behavior*. Transaction Books, 2007. 44
- [58] Inbal Ben-Ami Bartal, Jean Decety, and Peggy Mason. Empathy and pro-social behavior in rats. *Science*, 334(6061):1427–1430, 2011. 46
- [59] Stefan M Brudzynski. Communication of adult rats by ultrasonic vocalization: biological, sociobiological, and neuroscience approaches. *ILAR Journal*, 50(1):43–50, 2009. 46
- [60] D Caroline Blanchard, Robert J Blanchard, and Guy Griebel. Defensive responses to predator threat in the rat and mouse. *Current Protocols in Neuroscience*, pages 8–19, 2005. 46
- [61] Jaak Panksepp, Jeff Burgdorf, Margery C Beinfeld, Roger A Kroes, and Joseph R Moskal. Regional brain cholecystokinin changes as a function of friendly and aggressive social interactions in rats. *Brain research*, 1025(1):75–84, 2004. 46
- [62] Robert Assini, Yevgeniy B Sirotin, and Diego A Laplagne. Rapid triggering of vocalizations following social interactions. *Current Biology*, 23(22):R996–R997, 2013. 46
- [63] Stefan M Brudzynski and Dorota Ociepa. Ultrasonic vocalization of laboratory rats in response to handling and touch. *Physiology & behavior*, 52(4):655–660, 1992. 46
- [64] Darin J Knapp and Larissa A Pohorecky. An air-puff stimulus method for elicitation of ultrasonic vocalizations in rats. *Journal of neuroscience methods*, 62(1):1–5, 1995. 46
- [65] Connie Sánchez. Stress-induced vocalisation in adult animals. a valid model of anxiety? *European journal of pharmacology*, 463(1):133–143, 2003. 46

-
- [66] Eun Joo Kim, Earnest S Kim, Ellen Covey, and Jeansok J Kim. Social transmission of fear in rats: the role of 22-khz ultrasonic distress vocalization. *PLoS One*, 5(12):e15077, 2010. 46
- [67] Stefan M Brudzynski and Anna Pniak. Social contacts and production of 50-khz short ultrasonic calls in adult rats. *Journal of Comparative Psychology*, 116(1):73, 2002. 46, 47
- [68] Michal Bialy, Marek Rydz, and Leszek Kaczmarek. Precontact 50-khz vocalizations in male rats during acquisition of sexual experience. *Behavioral neuroscience*, 114(5):983, 2000. 46
- [69] Jaak Panksepp and Jeffrey Burgdorf. 50-khz chirping (laughter?) in response to conditioned and unconditioned tickle-induced reward in rats: effects of social housing and genetic variables. *Behavioural brain research*, 115(1):25–38, 2000. 46
- [70] Jeffrey Burgdorf, Brian Knutson, Jaak Panksepp, and Satoshi Ikemoto. Nucleus accumbens amphetamine microinjections unconditionally elicit 50-khz ultrasonic vocalizations in rats. *Behavioral neuroscience*, 115(4):940, 2001. 46, 48
- [71] Jeffrey Burgdorf, Brian Knutson, and Jaak Panksepp. Anticipation of rewarding electrical brain stimulation evokes ultrasonic vocalization in rats. *Behavioral neuroscience*, 114(2):320, 2000. 46
- [72] Jeffrey Burgdorf, Roger A Kroes, Joseph R Moskal, James G Pfaus, Stefan M Brudzynski, and Jaak Panksepp. Ultrasonic vocalizations of rats (*rattus norvegicus*) during mating, play, and aggression: Behavioral concomitants, relationship to reward, and self-administration of playback.. *Journal of comparative psychology*, 122(4):357, 2008. 47
- [73] Markus Wöhr and Rainer KW Schwarting. Ultrasonic communication in rats: can playback of 50-khz calls induce approach behavior? *PLoS One*, 2(12):e1365, 2007. 47
- [74] Markus Wöhr, Bart Houx, Rainer KW Schwarting, and Berry Spruijt. Effects of experience and context on 50-khz vocalizations in rats. *Physiology & behavior*, 93(4):766–776, 2008. 47

BIBLIOGRAFÍA

- [75] Stefan M Brudzynski. *Handbook of mammalian vocalization: an integrative neuroscience approach*, volume 19. Access Online via Elsevier, 2009. 47, 91, 93
- [76] Steffen R Hage and Uwe Jürgens. Localization of a vocal pattern generator in the pontine brainstem of the squirrel monkey. *European Journal of Neuroscience*, 23(3):840–844, 2006. 47
- [77] Uwe Jürgens. Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2):235–258, 2002. 47
- [78] C-L Lu and U Jürgens. Effects of chemical stimulation in the periaqueductal gray on vocalization in the squirrel monkey. *Brain research bulletin*, 32(2):143–151, 1993. 47
- [79] F Düsterhöft, U Häusler, and U Jürgens. Neuronal activity in the periaqueductal gray and bordering structures during vocal communication in the squirrel monkey. *Neuroscience*, 123(1):53–60, 2004. 48
- [80] Eva Dujardin and Uwe Jürgens. Afferents of vocalization-controlling periaqueductal regions in the squirrel monkey. *Brain research*, 1034(1):114–131, 2005. 48
- [81] Stefan M Brudzynski. Pharmacological and behavioral characteristics of 22khz alarm calls in rats. *Neuroscience & Biobehavioral Reviews*, 25(7):611–617, 2001. 48
- [82] Stefan M Brudzynski. Ultrasonic calls of rats as indicator variables of negative or positive states: acetylcholine–dopamine interaction and acoustic coding. *Behavioural brain research*, 182(2):261–273, 2007. 48
- [83] Briar Thompson, Kevin C Leonard, and Stefan M Brudzynski. Amphetamine-induced 50khz calls from rat nucleus accumbens: A quantitative mapping study and acoustic analysis. *Behavioural brain research*, 168(1):64–73, 2006. 48
- [84] D Sutton, C Larson, and RC Lindeman. Neocortical and limbic lesion effects on primate phonation. *Brain research*, 71(1):61–75, 1974. 48
- [85] Valerie Beller Domesick. Projections from the cingulate cortex in the rat. *Brain Research*, 12(2):296–320, 1969. 48

-
- [86] JT Cowles and LA Pennington. An improved conditioning technique for determining auditory acuity of the rat. *The Journal of Psychology*, 15(1):41–47, 1943. 48
- [87] Marvin J Herbert. An improved technique for studying the conditioned squeak reaction in hooded rats. *The Journal of general psychology*, 34(1):67–77, 1946. 48
- [88] OH Mowrer, Florence Palmer, and Marjorie D Sanger. Individual learning and “racial experience” in the rat, with special reference to vocalization. *The Pedagogical Seminary and Journal of Genetic Psychology*, 73(1):29–43, 1948. 49
- [89] Harbans Lal. Operant control of vocal responding in rats. *Psychonomic Science*, 8(1):35–36, 1967. 49
- [90] WI Welker. Analysis of sniffing of the albino rat. *Behaviour*, pages 223–244, 1964. 49
- [91] G. Bradski. *Dr. Dobb’s Journal of Software Tools*. 50
- [92] Timothy J Gardner and Marcelo O Magnasco. Sparse time-frequency representations. *Proceedings of the National Academy of Sciences*, 103(16):6094–6099, 2006. 54, 56
- [93] Sean A Fulop and Kelly Fitz. Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *The Journal of the Acoustical Society of America*, 119:360, 2006. 56
- [94] Kunihiro Kodera, Claude De Villedary, and Roger Gendrin. A new method for the numerical analysis of non-stationary signals. *Physics of the Earth and Planetary Interiors*, 12(2):142–150, 1976. 56
- [95] Meliza Daniel. fast multitaper conventional and reassignment spectrograms. <https://github.com/melizalab/libtfr>. 56
- [96] Andrey Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. 1971. 66

BIBLIOGRAFÍA

- [97] Kenneth P Burnham and David R Anderson. Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004. 79
- [98] Del D Thiessen, E Melanie, W Kittrell, and JM Graham. Biomechanics of ultrasound emissions in the mongolian gerbil, *Meriones unguiculatus*. *Behavioral and Neural Biology*, 29(4):415–429, 1980. 85
- [99] Mark S Blumberg. Rodent ultrasonic short calls: Locomotion, biomechanics, and communication. *Journal of Comparative Psychology*, 106(4):360, 1992. 85
- [100] Michael S Howe. *Acoustics of fluid-structure interactions*. Cambridge University Press, 1998. 90, 91
- [101] Michał Januszewski. Sailfish v 2013.1. <http://sailfish.us.edu.pl/>, 2011–2013. 94