

2021

ESTIMACIÓN DEL FLUJO DE AIRE GLOTAL MEDIANTE FILTRAJE Y SUAVIZAMIENTO DE KALMAN USANDO MODELOS DE BAJO ORDEN

MORALES BUGUEÑO, ARTURO RAÚL HUMBERTO

<https://hdl.handle.net/11673/52705>

Repositorio Digital USM, UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA

**ESTIMACIÓN DEL FLUJO DE AIRE GLOTOAL
MEDIANTE FILTRAJE Y SUAVIZAMIENTO DE
KALMAN USANDO MODELOS DE BAJO ORDEN**

Tesis de Grado presentada por

Arturo Raúl Humberto Morales Bugueño

como requisito parcial para optar al título de

Ingeniero Civil Electrónico

y al grado de

Magíster en Ciencias de la Ingeniería Electrónica

Profesor Guía

Dr. Juan Yuz Eissmann

Profesor Co-guía

Dr. Matías Zañartu Salas

Valparaíso, 2021.

TÍTULO DE LA TESIS:

**ESTIMACIÓN DEL FLUJO DE AIRE GLOTAL MEDIANTE
FILTRAJE Y SUAVIZAMIENTO DE KALMAN USANDO
MODELOS DE BAJO ORDEN**

AUTOR:

Arturo Raúl Humberto Morales Bugueño

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el título de Ingeniero Civil Electrónico y el grado de Magíster en Ciencias de la Ingeniería Electrónica de la Universidad Técnica Federico Santa María.

Dr. Juan Yuz E.

Dr. Matías Zañartu S.

Dr. Alejandro Weinstein O.

Valparaíso, Septiembre de 2021.

*Dedicados a mis padres Jacqueline y Arturo,
mi hermano Diego, mi abuela Yayita y
mi compañera incondicional Darinka*

AGRADECIMIENTOS

Este camino que he recorrido es un conjunto de eventos y personas que, “aleatoriamente”, se han cruzado en mi camino. Todas las vivencias que viví y todas las personas que conocí dieron como resultado el hecho de que yo esté hoy aquí. Por lo anterior, puede que me extienda más de lo normal, pero creo necesario agradecerles a todas las personas que, creo, me han dado ese Δ que me ha permitido llegar hasta aquí. Quien lea esto debe entender que para mí, el estudiar Ingeniería Civil Electrónica ya era un sueño que veía lejano. Hacerlo en una de las mejores universidades de ingeniería en Chile lo veía aún más lejano. Terminarlo con el grado de magíster, ya era algo impensado. Entonces, podría decirse que para mí, esto es algo así como un *error en un código*: algo que no debería ocurrir, pero que, por alguna razón, ocurre.

Me gustaría partir agradeciéndole a mi familia, porque son ellos los que me han apoyado en todos los momentos de mi vida. En específico, le agradezco a mi madre Jacqueline y mi padre Arturo por su amor incondicional, por siempre estar acompañándome a la distancia y por entregarme las herramientas para ser quien soy actualmente. A mi hermano Diego, el hermano que siempre quise, por ser un gran hermano y una mejor persona, y enseñarme a ser su hermano mayor. A mi *Yayita*, por amarme, criarme y cuidarme durante mis primeros 17 años de existencia. Y a las personas que ya no me acompañan físicamente pero que sé que desde algún lugar me dan su energía: mis abuelos paternos *Yujito* y *Tatita* y mi tío Raúl.

Es probable que este camino tan difícil lo hubiese sido aún más sin la compañía de cierta persona. Por lo anterior, me gustaría agradecerle a mi *Dari*, mi compañera incondicional que me ha acompañado durante casi 5 años. Agradezco tu paciencia infinita, pues sé que no es fácil estar con alguien que le apasiona el estudio y le gusta tomar más responsabilidades de las que puede. Créeme que tu amor y compañía han sido una fuerza revitalizadora que me ha amparado durante todo este duro proceso. Te amo infinito ♡.

Un/a profesor/a transforma vidas, negativa o positivamente. Por lo anterior, me gustaría agradecer a profesores/as que han sido vitales en mi proceso de aprendizaje y crecimiento como persona. Partiendo por mi escuela E-79, les agradezco profundamente a mis profesores/as Andrea Yapor, Pedro Saavedra, Luis Barraza y Marcos Cornejo, por confiar en mis capacidades y hacerme sentir especial e inteligente (aunque no lo fuera). De la enseñanza media, al profesor Luis Sepúlveda por su tremenda calidad humana y a la Sra. Adriana Barrientos por ayudarme desinteresadamente cuando lo requerí. De la universidad, debo confesar que me encontré con profesores/as inspiradores/as, sobre todo en el ámbito intelectual y humano. Personalmente me siento orgulloso del Departamento de Electrónica y su gente. En especial, me gustaría agradecerle a la profesora María José Escobar (mi tutora) que siempre tuvo una gran disposición para ayudarme y me brindó oportunidades para explorar nuevas áreas del conocimiento. Le agradezco también al profesor Matías

Zañartu, quien me ayudó en el transcurso de esta tesis entregándome información valiosa y cuyo trabajo (y alcance) es realmente admirable. Finalmente, me gustaría agradecerle profundamente al profesor Juan Yuz, que me permitió ser su ayudante, me aceptó como su alumno de magíster y me guió durante todo este proceso. Muchas gracias profesor por toda su ayuda, sobre todo por *enderezarme* cuando a veces mi mente se desviaba del camino, por aconsejarme en varios momentos y por ser un profesor que inspira y marca la diferencia.

También me gustaría agradecerle a los amigos/as que me han acompañado en distintos momentos de mi vida. A Iván Barrios y los demás amigos de los *Chaces III*, que me enseñaron el valor de la *amistad de barrio*. A mis mejores amigos, los cuales conocí en la E-79: Yimmy Chirino, Nicolás Ávila y Antonio Chaparro, a quienes les agradezco por tantos años de amistad, noches de juegos y series. A mis amigas del liceo: Daniela Molina, Karol Díaz y Katherine Lobos, por su compañía y amistad genuina. A mis amigos del DeLTA: Kevin Naranjo, Fabián Alarcón y Enrique García, por su amistad en tiempos difíciles. A mis amigos/as de la UA, en especial a Bastián Pizarro y Yicel Rojas, por su cariño y fe en mí. Y, obviamente, a mis amigos/as UTFSM, en especial a quienes conocí de *mechón*: Sebastián Neira, Benito Troncoso, Hernán Oyanadel, Héctor Ferreira, Rodrigo Muñoz, Ignacio Valenzuela y Fernanda Díaz, a mis amigos de trabajos: Sebastián Ubierno y Cedric Mulsow, al equipo *Stongol* que me dio la oportunidad de mantenerme activo en el fútbol y ganar campeonatos y a los amigos/as a quienes me acerqué en el último tiempo: Steev Gonzalez, Diego Allende, Yerald y Cabrera, Felipe Muñoz y Matias Licanqueo. Créanme que sin todos/as ustedes, hubiese sido tremendamente difícil aguantar la soledad alejado a más de 1000 km de distancia de mi familia.

Aunque pueda sonar extraño, me gustaría agradecerle a Don Federico Santa María. Las razones son varias, pero principalmente le agradezco por destinar su tremenda herencia a construir una universidad de ensueños, fundada para educar al *desvalido meritorio*, la cual me enamoró desde el primer día en que la vi y que me ha sabido dar probablemente los mejores momentos de mi vida. Es posible que sin su hermosa arquitectura o los magníficos atardeceres que se posaban junto al infinito mar, yo no hubiese llegado hasta este punto.

Finalmente, debo agradecer el soporte económico brindado por distintas instituciones que me permitieron estudiar este magíster con cierta tranquilidad económica. En específico, este trabajo fue financiado por la Agencia Nacional de Investigación y Desarrollo, ANID, BECAS/MAGÍSTER NACIONAL 22200156 y por las becas internas de la Universidad Técnica Federico Santa María. También, ha sido apoyado por el Centro Avanzado en Ingeniería Eléctrica y Electrónica, AC3E, Proyecto Basal FB0008, ANID y por el proyecto FONDECYT 1181090.

RESUMEN

Las patologías relacionadas con la voz son un problema de creciente importancia en materia de salud pública. Identificar las condiciones de abuso y/o mal uso del mecanismo vocal (es decir, la hiperfunción vocal) que, mantenidas en un período de tiempo, derivan en dichas patologías, es clave para poder prevenirlas y/o tratarlas. Métricas objetivas que han mostrado ser adecuadas para evaluar la presencia de hiperfunción vocal son las características aerodinámicas del flujo de aire que pasa por la glotis. En consecuencia, estimar el flujo de aire glotal resulta de gran apoyo en la tarea de prevenir y/o tratar trastornos comunes de la voz.

Si bien ya es posible obtener estimaciones del flujo de aire glotal en un ambiente de laboratorio, la literatura reciente se ha enfocado en realizarlo de manera ambulatoria y no invasiva. La evaluación ambulatoria permite caracterizar el comportamiento vocal asociado al uso de la voz durante las actividades diarias de los sujetos, lo que supone una evaluación de la hiperfunción vocal más confiable con respecto a la evaluación en laboratorio. Para ello, se ha propuesto el modelo basado en impedancia subglotal, el cual modela, en el dominio de la frecuencia, al sistema respiratorio junto a la piel y sus propiedades, para estimar el flujo de aire glotal a partir de las mediciones de un acelerómetro puesto sobre la piel del cuello.

En la presente tesis se trata el problema de estimar el flujo de aire glotal utilizando un enfoque Bayesiano aplicado sobre modelos en variables de estado de bajo orden. Este tipo de enfoques entrega una medida de la incertidumbre que existe en las estimaciones, al mismo tiempo que aminora el error producido al estimar el flujo de aire glotal con una vocal distinta a la utilizada para calibrar el modelo. Para realizar lo anterior, se obtiene una representación en variables de estado del modelo basado en impedancia subglotal, para luego utilizar un suavizador de Kalman y así estimar el flujo de aire glotal a partir de la aceleración que se produce sobre la piel del cuello.

La contribución principal de esta tesis es la de proveer estimaciones del flujo de aire glotal mediante el uso de un suavizador de Kalman aplicado sobre un modelo en variables de estado de bajo orden. Los resultados numéricos obtenidos son similares a los de enfoques existentes en la literatura. Sin embargo, el esquema propuesto involucra un costo computacional bastante menor que otros esquemas Bayesianos presentes en la literatura. A su vez, el enfoque Bayesiano provee una medida de la incertidumbre en las estimaciones, lo que constituye información valiosa en un ambiente ambulatorio.

Palabras claves

Identificación de sistemas, Filtraje de Kalman, Suavizamiento de Kalman, Cuerdas vocales, Hiperfunción vocal.

ABSTRACT

Voice-related pathologies are a problem of growing importance in public health. Identifying the conditions of abuse and/or misuse of the vocal mechanism (i.e. vocal hyperfunction) that lead to these pathologies, is key to be able to prevent and/or treat them. Objective metrics that have shown to be adequate to assess the presence of vocal hyperfunction are the aerodynamic characteristics of the airflow through the glottis. In conclusion, estimating the glottal airflow may help in the task of preventing and/or treating common voice disorders.

Although it is already possible to obtain estimates of the glottal airflow in a laboratory environment, the state-of-art is focused on ambulatory and non-invasive monitoring. Ambulatory assessment allows to characterize the vocal behavior associated with the voice during the subjects' daily activities, which suppose a more reliable assessment of the vocal hyperfunction with respect to laboratory assessment. With this aim, the subglottal impedance-based model has been proposed in the frequency domain, to model the respiratory system together the skin and its properties. This model has been used to estimate the glottal airflow from measurements of an accelerometer placed on the neck skin.

This thesis deals with the problem of estimating the glottal airflow using a Bayesian approach based on low order state-space models. Specifically, the proposed approach yields a measure of the uncertainty in the estimates, at the same time that reduces the error produced when the glottal airflow is estimated with a different vowel than the one used to calibrate the model. To carry out the above, a state-space representation for the subglottal impedance-based model is obtained, to then use a Kalman smoother and, in this way, to estimate the glottal airflow from the acceleration produced on the neck skin.

The main contribution in this thesis is to provide estimates of the glottal airflow through the use of a Kalman smoother based on a low order state-space model. The obtained numerical results are similar to current approaches presented in the literature. The proposed scheme, however, involves a lower computational cost, at the same time that it provides a measure of the uncertainty in the estimates, which is equivalent to valuable information in an ambulatory environment.

Keywords

System identification, Kalman filtering, Kalman smoothing, Vocal folds, Vocal hyperfunction.

ÍNDICE

AGRADECIMIENTOS	I
RESUMEN	III
ABSTRACT	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	X
1. INTRODUCCIÓN	1
1.1. Estado del arte	3
1.2. Definición del problema	4
1.2.1. Representación del modelo en variables de estado	5
1.2.2. Estimación del flujo de aire glotal	6
1.3. Principales contribuciones	6
1.4. Publicación asociada	6
1.5. Estructura del documento	6
2. PRELIMINARES	8
2.1. Producción del habla	8
2.2. Modelo glotal	11
2.2.1. Modelo de Butterworth	12
2.2.2. Modelo Rosenberg C	12
2.3. Características aerodinámicas del flujo de aire glotal	13
2.4. Identificación de sistemas	16
2.4.1. Teoría de realizaciones	19
2.4.2. Métodos de subespacios	21
2.4.3. Métodos de predicción del error	26
2.5. Reducción de modelos	28
2.6. Filtraje y suavizamiento Bayesiano	31
2.6.1. Filtraje Bayesiano	34
2.6.2. Suavizamiento Bayesiano	35

3. MODELO BASADO EN IMPEDANCIA SUBGLOTAL DE ORDEN REDUCIDO	37
3.1. Modelo basado en impedancia subglotal	37
3.2. Representación en variables de estado del modelo IB subglotal	45
3.2.1. Reducción de orden	46
3.2.2. Identificación en el dominio del tiempo	50
3.2.3. Identificación en el dominio de la frecuencia	55
4. FILTRAJE INVERSO MEDIANTE SUAVIZADOR DE KALMAN	58
4.1. Filtraje inverso	58
4.2. Filtraje inverso basado en impedancia subglotal	62
4.3. Filtraje inverso mediante filtraje de Kalman de alto orden	64
4.4. Filtraje inverso mediante suavizamiento de Kalman de bajo orden	73
4.4.1. Modelando la entrada como ruido blanco (WT)	73
4.4.2. Modelando la entrada como Random Walk (RW)	75
4.4.3. Modelando la entrada con el modelo de Butterworth (BW)	78
4.5. Medida de la incertidumbre	80
5. RESULTADOS NUMÉRICOS	82
5.1. Elección de parámetros	82
5.2. Flujo de aire glotal	85
5.3. Características aerodinámicas	86
5.4. Tiempos de cómputo	95
6. CONCLUSIONES	96
6.1. Trabajo futuro	98
A. DISTRIBUCIÓN GAUSSIANA	99
BIBLIOGRAFÍA	101

ÍNDICE DE FIGURAS

1.1. Acelerómetro puesto sobre la piel del cuello.	4
1.2. Esquema del problema a solucionar.	5
2.1. Vista de la sección transversal del mecanismo vocal.	9
2.2. Cuerdas vocales y glotis.	10
2.3. Modelo lineal de la producción del habla.	11
2.4. Modelo lineal de la producción del habla (Dominio z).	11
2.5. Modelo de Butterworth	13
2.6. Modelo Rosenberg C.	14
2.7. Características aerodinámicas del flujo de aire glotal y su derivada.	15
2.8. Sistema general con entrada, salida y ruido.	17
2.9. Diagrama de bloques usando funciones de transferencia.	17
3.1. Sistema subglotal modelado con tubos cilíndricos.	38
3.2. Representación del elemento básico usado en el modelo acústico del sistema subglotal.	39
3.3. Representación del modelo del tracto subglotal junto a la inclusión de Z_{skin}	40
3.4. Representación fisiológica del sistema subglotal.	41
3.5. Magnitud y fase de la respuesta en frecuencia del modelo basado en impedancia subglotal.	44
3.6. Respuesta a impulso del modelo basado en impedancia subglotal	44
3.7. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado construido a partir de una versión trasladada de h_t	48
3.8. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado construido a partir de la versión causal \tilde{h}_t de h_t	49
3.9. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido obtenido mediante truncamiento balanceado construido a partir de la respuesta a impulso causal.	51
3.10. Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido por reducción de modelo mediante truncamiento balanceado para modelos de distintos órdenes y considerando diferentes sujetos.	52

3.11. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido obtenido mediante identificación en el dominio del tiempo.	53
3.12. Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido mediante identificación en el dominio del tiempo para modelos de distintos órdenes de sistema y considerando diferentes sujetos.	54
3.13. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido obtenido mediante identificación en el dominio de la frecuencia.	56
3.14. Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido mediante identificación en el dominio de la frecuencia para modelos de distintos órdenes y considerando diferentes sujetos.	57
4.1. Modelo de la teoría fuente-filtro.	59
4.2. Filtraje inverso para la obtención del flujo de aire glotal.	59
4.3. Filtraje inverso a partir de la señal del habla.	60
4.4. Filtraje inverso a partir del flujo de aire oral.	60
4.5. Neumotacógrafo con ventilación circunferencial, más conocido como máscara de Rothenberg.	61
4.6. Flujo de aire glotal obtenido mediante el filtraje inverso del flujo de aire oral.	61
4.7. Modelo basado en impedancia subglotal.	62
4.8. Filtraje inverso basado en impedancia subglotal.	63
4.9. Flujo de aire glotal junto a su derivada obtenidos mediante IBIF subglotal.	65
4.10. Flujo de aire glotal junto a su derivada obtenidos mediante IBIF subglotal a partir de la vocal /i/.	66
4.11. Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado obtenido a partir de la respuesta a impulso truncada en el centro con una ventana de Hann.	69
4.12. Flujo de aire glotal junto a su derivada obtenidos mediante filtraje de Kalman.	71
4.13. Flujo de aire glotal junto a su derivada obtenidos mediante filtraje de Kalman a partir de la vocal /i/.	72
4.14. Flujo de aire glotal junto a su derivada obtenidos mediante suavizamiento de Kalman y el modelamiento del flujo de aire glotal como ruido blanco Gaussiano.	76
4.15. Flujo de aire glotal junto a su derivada obtenidos mediante suavizamiento de Kalman y el modelamiento del flujo de aire glotal como <i>Random Walk</i>	77
4.16. Flujo de aire glotal junto a su derivada obtenidos mediante suavizamiento de Kalman y el modelamiento del flujo de aire glotal usando el modelo de Butterworth.	79
4.17. Estimación del flujo de aire glotal obtenido mediante suavizamiento de Kalman usando el modelo de ruido blanco, junto a una banda que refleja la varianza del error de estimación.	81
5.1. Ejemplo de datos utilizados. Flujo de aire glotal obtenido a partir del flujo de aire oral para el sujeto FN001 utilizando la vocal /a/.	84

5.2. Valores de WMAE para distintas combinaciones (σ_w^2, σ_v^2) en el caso del sujeto FN001 al pronunciar la vocal /a/ y considerando el modelo WT de orden 15.	84
5.3. Gráfica de ACFL obtenido mediante la señal de referencia y mediante suavizamiento de Kalman usando el modelo de ruido blanco.	92

ÍNDICE DE TABLAS

2.1. Parámetros utilizados por el modelo de Rosenberg C.	13
3.1. Parámetros Q de los sujetos analizados en este trabajo.	45
5.1. Parámetros (σ_w^2, σ_v^2) del suavizador de Kalman elegidos en base al menor error absoluto medio ponderado obtenido para la vocal /a/.	85
5.2. Raíz del error cuadrático medio obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /a/.	87
5.3. Raíz del error cuadrático medio normalizado obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /a/.	87
5.4. Raíz del error cuadrático medio obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /i/.	88
5.5. Raíz del error cuadrático medio normalizado obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /i/.	88
5.6. Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /a/ (1 de 2).	90
5.7. Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /a/ (2 de 2).	90
5.8. Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /i/ (1 de 2).	91
5.9. Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /i/ (2 de 2).	91

5.10. Error relativo promedio al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada ventana de tiempo, considerando la vocal /a/.	93
5.11. Error relativo promedio al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada ventana de tiempo, considerando la vocal /i/.	93
5.12. Valores R^2 obtenidos al realizar una regresión lineal entre las características estimadas y las características de referencia, para cada ventana de tiempo, considerando la vocal /a/.	94
5.13. Valores R^2 obtenidos al realizar una regresión lineal entre las características estimadas y las características de referencia, para cada ventana de tiempo, considerando la vocal /i/.	94
5.14. Tiempos de cómputo necesarios para obtener estimaciones del flujo de aire glotal considerando diferentes enfoques.	95

INTRODUCCIÓN

Un problema de salud cuya importancia ha aumentado en nuestra sociedad en el último tiempo son aquellos relacionados con la voz, los cuales han adquirido mayor relevancia debido a diversos estudios que se han realizado para estimar la prevalencia que existe en la población de desarrollar algún trastorno en ella. En [1, 2] se encontró que un 30 % de la población general de Estados Unidos ha experimentado un trastorno de la voz en algún momento de su vida, mientras que en [3] se estimó que 1 de cada 13 estadounidenses (esto es, un 7.7 %) experimenta un problema relacionado con la voz anualmente. Similarmente, en Chile una de las enfermedades profesionales más frecuentes, con un 32 % de prevalencia sobre el total de patologías, corresponde a las disfonías que, de hecho, son un trastorno común de la voz [4].

Esto claramente aumenta si analizamos a profesionales que utilizan la voz como herramienta de trabajo, donde es importante diferenciar a aquellos que la utilizan profesionalmente y aquellos que la utilizan ocupacionalmente: los primeros, como cantantes o actores/actrices, son profesionales entrenados para utilizar la voz eficientemente, mientras que los segundos, como profesores/as o telefonistas, son aquellos que la utilizan como herramienta de trabajo pero que no han recibido un entrenamiento específico en su uso [5]. En el caso de profesionales ocupacionales, los que han sido más estudiados son los profesores, donde por ejemplo, en Estados Unidos, se ha encontrado que estos profesionales tienen un 57.7 % de prevalencia de tener un problema en la voz a lo largo de sus vidas contra sólo un 28.8 % de prevalencia que poseen otros tipos de trabajadores [2]. En Chile, por su parte, se ha encontrado que existe un 75.5 % de prevalencia de desarrollar algún trastorno de la voz siendo profesor, lo que puede ser explicado por las malas condiciones de infraestructura en los establecimientos educacionales, la excesiva cantidad de alumnos por curso, entre otros factores [4, 6].

Muchos de estos trastornos pueden aparecer por un mal uso de la voz prolongado en el tiempo, por lo que sería de suma importancia el poder detectar cuándo se está haciendo un mal uso de la voz y así tomar acciones al respecto, ayudando en la prevención de dichos trastornos. Tal es la importancia del concepto de *mal uso de la voz* que se acuñó el término *hiperfunción vocal (VH)* para referirse a “*las condiciones de abuso y/o mal uso del mecanismo vocal debido a las excesivas y/o desbalanceadas fuerzas musculares*” [7]. Podemos

distinguir dos tipos de VH: la VH no fonotraumática (NPVH) y la VH fonotraumática (PVH) [8]. Así, se ha identificado que si un individuo presenta hiperfunción vocal durante un tiempo prolongado, entonces es probable que desarrolle una patología en las cuerdas vocales [7, 8].

Por lo anterior, es importante determinar cuándo un sujeto presenta hiperfunción vocal, para lo cual es necesario contar con una manera objetiva de evaluar esta condición. Así, en [7] se hizo el primer reporte donde se le intenta dar una mirada objetiva a la hiperfunción vocal mediante la evaluación de métricas que permitan cuantificar dicha condición.

Dentro de las métricas objetivas que se han utilizado en la literatura, existen dos a las que se les ha prestado mayor atención: las aerodinámicas [7] y las acústicas [9]. Sin embargo, varios estudios [10–13] han mostrado que las medidas aerodinámicas identifican de mejor manera la hiperfunción vocal junto al tipo (NPVH ó PVH) en comparación con las medidas acústicas.

Dichas medidas o parámetros aerodinámicos son obtenidos a partir del flujo de aire que pasa por la glotis (junto a su derivada), por lo que es de suma importancia obtener dicha señal (también conocida como *velocidad volumétrica glotal* o GVV). Si bien el flujo de aire glotal ya puede ser obtenido en un ambiente de laboratorio [14], la literatura actual está enfocada en estimar dicha señal de manera ambulatoria y no invasiva, para así poder monitorear el uso de la voz y, por tanto, la existencia de hiperfunción vocal, durante las actividades diarias de los pacientes [15, 16]. Esto representa una ventaja por sobre la evaluación de la voz en un ambiente de laboratorio donde podría suceder que, en el momento de la evaluación, el paciente no muestre signos de hiperfunción vocal [17, 18].

Como se ha indicado, el flujo de aire glotal juega un rol importante en la evaluación objetiva y ambulatoria de la hiperfunción vocal. Es importante señalar que dicho flujo es una cantidad física que varía con el tiempo o, dicho de otra manera, es una *señal*. Así, es con la ayuda de la teoría de *señales y sistemas* [19, 20] que podemos abstraer los conceptos físicos-biológicos anteriores y utilizar herramientas como las transformadas Zeta o de Fourier, para analizar y procesar señales de interés.

Al mismo tiempo, mediante el *modelamiento de sistemas* es posible describir matemáticamente las relaciones entre señales. En nuestro caso particular, el flujo de aire que pasa por la glotis no es medible directamente, pero provoca que la piel del cuello vibre. Esta vibración sí es medible directamente y, además, de forma ambulatoria. Así, entre estas dos señales existe una relación que puede ser modelada y, por tanto, analizada por las herramientas matemáticas que nos entrega la teoría de señales y sistemas.

Con lo anterior, se abre la posibilidad de estimar el flujo de aire glotal mediante un enfoque Bayesiano. Este enfoque permite estimar señales en base a un modelo que relacione la señal a estimar y la señal medible. Es importante mencionar que dicho enfoque Bayesiano implica trabajar con matrices cuyos tamaños dependen del orden del modelo, por lo que una restricción práctica es que el modelo obtenido debe ser de un orden tal que su implementación en un dispositivo real sea factible.

Así, en este trabajo de tesis se busca mejorar, desde una perspectiva computacional, los trabajos anteriores que han hecho uso del modelo que relaciona al flujo de aire glotal con la aceleración de la piel del cuello en el dominio de la frecuencia. Con dicho modelamiento en frecuencia es posible construir un modelo en variables de estado y posteriormente aplicar técnicas de filtraje y suavizamiento Bayesiano con el fin de *invertir* la relación entre ambas señales y obtener el flujo de aire glotal mediante la medición de la aceleración sobre la piel del cuello.

1.1. Estado del arte

En la literatura se pueden encontrar diversos artículos sobre la hiperfunción vocal, su diagnóstico objetivo y las técnicas empleadas para obtener señales de las cuales sea posible extraer métricas que la cuantifiquen objetivamente. Uno de los primeros trabajos donde se da una descripción objetiva de la hiperfunción vocal es [7] (mientras que en [8] se realiza una actualización a dicho trabajo). En dicho trabajo además se hipotetiza sobre los dos tipos de hiperfunción vocal y se proponen métricas objetivas para identificarla.

De entre todas las métricas que podrían servir para evaluar la hiperfunción vocal, hay dos tipos que han mostrado una capacidad de discriminación: las aerodinámicas [7] y las acústicas [9]. Sin embargo, en [10–13, 17] se muestran estudios que indican que utilizando métricas aerodinámicas y, en específico, el flujo de aire que pasa por la glotis, es posible distinguir objetivamente a pacientes con hiperfunción vocal de algún tipo (PVH ó NPVH) de sujetos sanos, con una mayor fidelidad que utilizando, por ejemplo, métricas acústicas como el nivel de presión sonora (*Sound Pressure Level*, SPL) o la frecuencia fundamental. Previo a [7], las métricas aerodinámicas obtenidas del flujo de aire glotal parecían tener una alta relevancia en el diagnóstico de la hiperfunción vocal. Por ejemplo, en [14] se presenta una técnica de filtraje inverso para obtener el flujo de aire glotal a partir de la medición del flujo de aire oral (también conocido como *oral volume velocity* u OVV) obtenidas de un neumotacógrafo con ventilación circunferencial, más conocido como máscara de Rothenberg. Esta técnica es la que ha entregado las estimaciones más confiables de dicha señal, por lo que en muchos estudios es utilizada como señal de referencia o *ground truth*.

Con lo dicho anteriormente, es claro que la estimación del flujo de aire glotal para la obtención de métricas objetivas es clave en la evaluación y tratamiento de la hiperfunción vocal. Más importante aún es realizar lo anterior de manera ambulatoria y no invasiva, de manera de obtener resultados más fiables con respecto a la existencia de la hiperfunción vocal en comparación con realizarlo en un ambiente de laboratorio.

Uno de los enfoques ambulatorios y no invasivos que ha dado resultados bastantes prometedores en la evaluación de la hiperfunción vocal [17, 21] es el denominado *filtraje inverso basado en impedancia (IBIF) subglotal* [22]. En este enfoque se construye un modelo en frecuencia denominado *modelo basado en impedancia (IB) subglotal*, el cual relaciona el flujo de aire que pasa por la glotis con la aceleración sobre la piel del cuello, para luego obtener estimaciones del flujo de aire glotal utilizando mediciones adquiridas por un acelerómetro (Fig. 1.1). Lo anterior se realiza mediante el uso de la transformada discreta de Fourier (DFT) y la inversión en el dominio de la frecuencia. Estimar el flujo de aire glotal de dicha forma tiene el inconveniente de no proveer una medida sobre la incertidumbre en las estimaciones, lo que puede ser una desventaja en una situación ambulatoria donde no se tiene acceso a señales de referencia. Otra desventaja de este enfoque es la obtención de los parámetros utilizados en la construcción del modelo en frecuencia, los cuales son específicos para cada paciente y son obtenidos en un ambiente de laboratorio. Por lo tanto, podría suceder que éstos cambien en una situación ambulatoria y, por ende, el modelo resultante sólo entregue buenas estimaciones en un ambiente *controlado*, esto es, con el sujeto en estudio inmóvil o manteniendo el cuello sólo en cierta posición.

Otro enfoque ambulatorio y no invasivo se propone en [23], donde se intentan mejorar las desventajas del método mencionado anteriormente mediante el planteamiento del problema usando filtraje Bayesiano. Así, se utiliza el modelo en frecuencia obtenido en [22] para construir un modelo en variables de estado y de este modo aplicar un filtro de Kalman para estimar el flujo de aire glotal mediante el uso de las mediciones del acelerómetro. La ventaja



Figura 1.1: Acelerómetro puesto sobre la piel del cuello [16].

de este esquema es que el filtro de Kalman, además de estimar la señal de interés, entrega una medida de la incertidumbre de dichas estimaciones. Además, la forma en que puede ser planteado el filtro de Kalman permite incorporar la incertidumbre o desconocimiento de los parámetros del modelo real. A pesar de las ventajas de este enfoque, la forma en que fue planteado trajo consigo un alto costo computacional debido a que el sistema en variables de estado utilizado para aplicar el filtro de Kalman es de alto orden, haciendo que la implementación de esta solución resulte compleja desde una perspectiva computacional.

Debido a lo anterior y a los prometedores resultados de la evaluación ambulatoria de la hiperfunción vocal, es válido adentrarse en la mejora computacional de los métodos existentes. Así, estimar el flujo de aire glotal mediante un enfoque de filtraje y suavizamiento de Kalman de bajo orden y, por tanto, computacionalmente poco costoso, resulta ser un aporte, que no ha sido estudiado, en la evaluación de la hiperfunción vocal.

1.2. Definición del problema

El problema a considerar en esta tesis es el de estimar el flujo de aire que pasa por la glotis mediante el uso de las mediciones de un acelerómetro y un suavizador de Kalman de bajo orden. Así, para comprender el problema considere el diagrama de bloques mostrado en la Fig. 1.2, el cual será explicado en mayor profundidad en los capítulos 2 y 4. Aquí basta con entender que la teoría lineal de la producción del habla modela a ésta como la salida de un conjunto de filtros lineales excitados por el flujo de aire glotal. A su vez, este flujo puede ser modelado como la salida de un filtro excitado por una señal periódica, la cual no es medible.

Como se ha dicho anteriormente, la obtención del flujo de aire glotal ya puede ser obtenida mediante el flujo de aire oral y la inversión del filtro *Tracto Vocal* [14], sin embargo dicho flujo sólo puede ser obtenido en un ambiente de laboratorio. Así, para obtener el flujo de aire glotal de una forma ambulatoria, en [22] se construye un modelo basado en impedancia que relaciona dicho flujo con la señal de un acelerómetro para posteriormente invertir dicha relación en el dominio de la frecuencia y obtener el flujo de aire glotal a partir de las mediciones del acelerómetro. Luego, como nos interesa realizar lo anterior utilizando un suavizador de Kalman, es necesario obtener una representación en variables de estado de

bajo orden, con el fin que sea implementable.

Por lo anterior, el problema a considerar en este trabajo de tesis puede dividirse en dos subproblemas, donde el primero trata sobre obtener una representación en variables de estado de bajo orden que describa al modelo que relaciona el flujo de aire glotal con la aceleración sobre la piel del cuello y el segundo trata de estimar, usando un suavizador de Kalman, dicho flujo de aire a partir de las mediciones de un acelerómetro. Es claro que el primer subproblema es clave en este trabajo, pues para poder hacer uso de un suavizador de Kalman es necesario conocer el modelo cuya salida es la señal medida y cuyo vector de estado contiene explícita o implícitamente la señal a estimar. A continuación, en las siguientes subsecciones se describe brevemente cada subproblema.

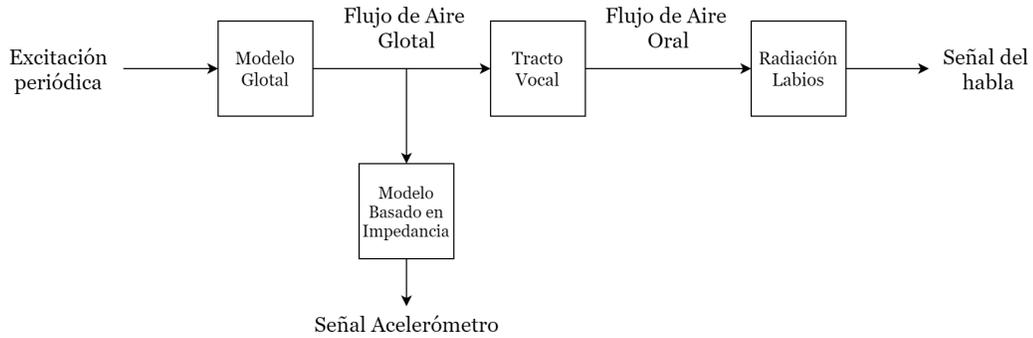


Figura 1.2: Esquema del problema a solucionar. Los bloques representan sistemas, mientras que los términos intermedios representan señales.

1.2.1. Representación del modelo en variables de estado

Para poder utilizar un suavizador de Kalman, es imperativo conocer el modelo que relaciona las señales medidas y las señales a estimar. En específico, consideremos un modelo lineal e invariante en el tiempo (LTI) representado en variables de estado y dado por las siguientes ecuaciones:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \quad (1.1)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t \quad (1.2)$$

donde $\mathbf{x}_t \in \mathbb{R}^n$ es el vector de estados del sistema, $\mathbf{u}_t \in \mathbb{R}^p$ es el vector de entradas del sistema, $\mathbf{y}_t \in \mathbb{R}^m$ es el vector de salidas del sistema, $\mathbf{w}_t \in \mathbb{R}^n$ es el vector de ruidos de proceso y $\mathbf{v}_t \in \mathbb{R}^m$ es el vector de ruidos de salida, mientras que $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, $\mathbf{C} \in \mathbb{R}^{m \times n}$ y $\mathbf{D} \in \mathbb{R}^{m \times p}$ son las matrices de estado, de entrada, de salida y de paso directo, respectivamente. Además, los ruidos \mathbf{w}_t y \mathbf{v}_t los supondremos Gaussianos de media 0 con matrices de covarianza \mathbf{Q} y \mathbf{R} , respectivamente, y matriz de covarianza cruzada \mathbf{S} .

Como el suavizador y el filtro de Kalman estiman el vector \mathbf{x}_t utilizando como datos las mediciones de \mathbf{y}_t , en nuestro caso debemos construir un modelo que relacione al flujo de aire glotal, u_t^g , con la señal del acelerómetro, y_t^a , de tal manera que \mathbf{x}_t incluya implícita o explícitamente a u_t^g , \mathbf{y}_t sea igual a y_t^a y donde además debemos notar que en nuestro caso particular no existe una señal de entrada \mathbf{u}_t medible. Por lo anterior, este subproblema se podría resumir como encontrar las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{D} del menor orden posible que

describan, sin perder demasiada precisión, la relación entre el flujo de aire glotal y la señal del acelerómetro, junto con encontrar las matrices de covarianza \mathbf{Q} , \mathbf{R} y \mathbf{S} .

1.2.2. Estimación del flujo de aire glotal

Posterior a tener un modelo como el mostrado en las ec. (1.1) - (1.2), es posible utilizar un enfoque Bayesiano para estimar el estado \mathbf{x}_t a partir de las mediciones de \mathbf{y}_t . Debido a que las estimaciones no se necesitan de una forma *online*, es posible utilizar un suavizador Bayesiano, el cual hace uso de mediciones de \mathbf{y}_t obtenidas antes y después del instante en el que se desea estimar \mathbf{x}_t , lo que provee mejores estimaciones de \mathbf{x}_t junto a menores covarianzas del error de estimación en comparación a un filtro Bayesiano.

En nuestro caso, al tener un modelo lineal y al suponer los ruidos de proceso y de salida como Gaussianos, el suavizador Bayesiano óptimo se conoce como suavizador RTS o de Kalman, el cual es un suavizador de intervalo fijo. Así, este segundo subproblema trata sobre implementar el suavizador RTS y, de esta manera, obtener estimaciones del flujo de aire glotal, el cual debe estar incluido implícita o explícitamente en el vector de estados \mathbf{x}_t como se mencionó anteriormente.

1.3. Principales contribuciones

Las principales contribuciones de esta tesis son:

- Se obtiene un modelo descrito en variables de estado de bajo orden que relaciona el flujo de aire glotal con la aceleración sobre la piel del cuello. Este modelo es obtenido utilizando tres técnicas distintas: *métodos de predicción del error* (PEM, por sus siglas en inglés) con datos en el dominio del tiempo, PEM con datos en el dominio de la frecuencia y reducción de modelo mediante truncamiento balanceado.
- Se proponen distintos modelos descritos en variables de estado de bajo orden donde la señal medida es la aceleración sobre la piel del cuello y el flujo de aire glotal está incluido implícitamente en el vector de estados.
- Se implementa un suavizador de Kalman para estimar el flujo de aire glotal a partir del modelo obtenido y de las mediciones de un acelerómetro.
- Se estima el flujo de aire glotal de manera ambulatoria y no invasiva utilizando una solución computacionalmente poco costosa, la cual, además, entrega una medida de la incertidumbre en el error de las estimaciones.

1.4. Publicación asociada

- [Arturo Morales](#) and [J.I. Yuz](#). **Reduced order modeling for glottal airflow estimation using a Kalman smoother**. In IEEE IFAC International Conference on Automation ICA ACCA, 22-26 March 2021.

1.5. Estructura del documento

La estructura del presente documento se describe a continuación:

- **Capítulo 2:** En este capítulo se presentan las herramientas preliminares que son utilizadas a lo largo de los siguientes capítulos.

- **Capítulo 3:** En este capítulo se presenta el modelo basado en impedancia (IB) subglotal y se discute cómo obtener una representación en variables de estado de bajo orden. Se analizan tres enfoques diferentes: usando métodos de identificación en el dominio del tiempo, en el dominio de la frecuencia y métodos de reducción de orden.
- **Capítulo 4:** En este capítulo se utiliza la representación en variables de estado del modelo IB subglotal para estimar el flujo de aire glotal a partir de las mediciones de un acelerómetro puesto sobre la piel del cuello. En específico, se proponen diferentes métodos para introducir el flujo de aire glotal en el vector de estados, para luego implementar un suavizador de Kalman con el fin de obtener estimaciones de dicho flujo junto a la varianza del error de estimación.
- **Capítulo 5:** En este capítulo se exponen resultados numéricos. En específico, se analiza el desempeño del enfoque aquí propuesto mediante la comparación con otros métodos existentes en la literatura, tanto al momento de obtener métricas objetivas como en la complejidad computacional.
- **Capítulo 6:** Finalmente, en este capítulo se entregan conclusiones de la presente tesis y se discute el alcance de futuros trabajos.

PRELIMINARES

En el presente capítulo se describe el proceso de la producción del habla, junto a un modelo lineal usado para comprenderlo. Posteriormente, se muestran algunos modelos glotales utilizados para describir la señal de interés: el flujo de aire glotal. Luego, se hace una revisión general sobre los conceptos de identificación de sistemas, reducción de modelos y estimación Bayesiana que serán utilizados en los capítulos posteriores.

2.1. Producción del habla

El habla es un proceso complejo que ha servido al ser humano para comunicarse con sus pares y que tardó millones de años de evolución para llegar a convertirse en la herramienta que aprendemos en nuestros primeros años de vida [24]. Estudiar este complejo proceso conlleva un trabajo multidisciplinario, donde científicos de diversas áreas como la paleoantropología, primatología, ciencias del habla, anatomía, acústica, entre otras, han intentado responder a diversas preguntas, desde distintas perspectivas, para arrojar luz en su entendimiento. Sin embargo, para comprender el trabajo desarrollado en la presente tesis, basta con entender lo básico sobre dicho proceso, además de conocer las señales importantes involucradas en él (algunas de las cuales nos son de principal interés). De esta forma podemos luego utilizar nuestros conocimientos de señales y sistemas con el fin de abstraer las ideas subyacentes a él y poder analizarlo.

Para comenzar, debemos entender que la señal del habla, que es la que *sale* de nuestra boca cuando hablamos, es una *onda de presión acústica* que otras personas pueden percibir como sonido gracias a sus oídos. Esta onda de presión es generada mediante el mecanismo vocal que se muestra en la Fig. 2.1, el cual incluye a:

- Los pulmones junto a la caja torácica como fuente de aire para excitar el tracto vocal y como fuente de presión para expulsar el aire desde los pulmones.
- La tráquea como conductor del aire desde los pulmones hacia las cuerdas vocales y el tracto vocal.
- Las cuerdas vocales (Fig. 2.2), las cuales son pliegues que vibran debido al flujo de aire que pasa por su abertura o *glotis*.

- El tracto vocal, que consiste en la faringe, la cavidad bucal (que, a su vez, incluye la lengua, labios, mandíbula y boca) y la cavidad nasal y cuya función es determinar el sonido producido.

Así, el mecanismo de producción del habla para los sonidos con voz¹ (o *voiced sounds*) como los producidos al decir una vocal funciona de la siguiente manera [25,26]:

- El aire entra a los pulmones mediante la respiración normal.
- A medida que el aire es expulsado desde los pulmones a través de la tráquea, las cuerdas vocales tensas dentro de la laringe se hacen vibrar debido a las variaciones de la presión del aire en la abertura glotal.
- El flujo de aire se transforma en pulsos cuasi-periódicos mediante la apertura y cierre del orificio glotal.
- Estos pulsos pasan a través de la faringe, la cavidad bucal y la cavidad nasal. Finalmente, las posiciones de la mandíbula, lengua, labios y boca determinan el sonido producido.

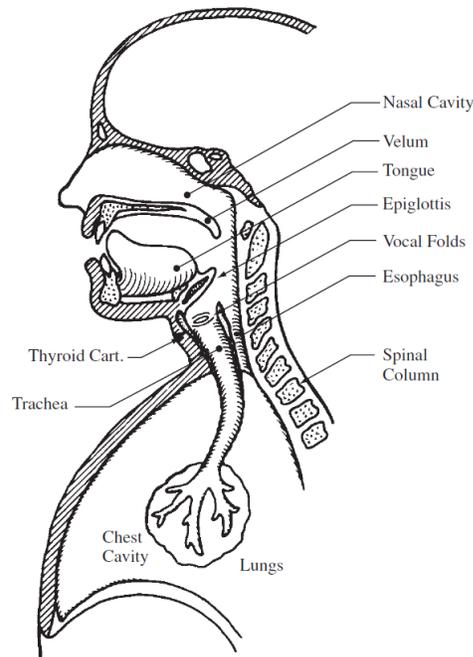


Figura 2.1: Vista de la sección transversal del mecanismo vocal [26].

En este punto es necesario especificar que en este trabajo de tesis la señal que nos interesa estimar utilizando modelos de bajo orden es el flujo de aire glotal, también llamado *glottal volume velocity (GVV)*. Este es el flujo de aire que pasa por la glotis (ver Fig. 2.2) y, como

¹Aunque también existen los sonidos sin voz o *unvoiced sounds*, en la presente tesis sólo se trabajará con sonidos con voz.

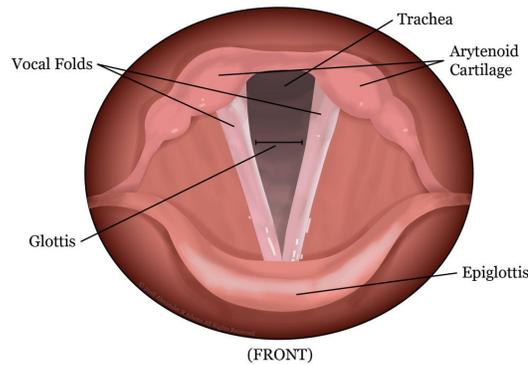


Figura 2.2: Las cuerdas vocales (en inglés *vocal folds*) son pliegues de tejido en la garganta que son clave para crear sonidos a través de la vocalización. La glotis (en inglés *glottis*) es la abertura entre las cuerdas vocales.

veremos a continuación, puede ser visto como la fuente de excitación que, al pasar por el tracto vocal, da como resultado la señal del habla.

Una vez comprendido cualitativamente el proceso de la producción del habla, es posible utilizar modelos matemáticos para describirlo de manera cuantitativa y así obtener las cantidades físicas de interés que nos permiten entender dicho proceso de una manera más objetiva. Un modelo matemático puede ser tan complejo o simple como se quiera, donde en general se esperaría que un modelo complejo retrate de manera más fiel la realidad que un modelo simple, a cambio de una mayor dificultad al tratar con él. A pesar de lo anterior, una clase de modelo que ha sido ampliamente estudiada son los modelos lineales, pues poseen características que los hacen simples al momento de trabajar, además de contar con herramientas poderosas para su análisis.

Es por la simplicidad y potencia de los modelos lineales que G. Fant, en [27], desarrolló el *modelo lineal de la producción del habla*. Este modelo puede ser entendido utilizando el diagrama de bloques de la Fig. 2.3 [25,26], donde la señal del habla es la salida de un conjunto de filtros en serie y excitados por una excitación periódica, la cual, en el caso de sonidos con voz² es un tren de impulsos de período T (llamado *pitch*). La excitación periódica entra al primer filtro llamado *Modelo Glotal*, el cual intenta modelar el flujo de aire que pasa por la glotis, y entrega como salida la señal de interés GVV. Luego, el flujo de aire glotal excita al filtro llamado *Tracto Vocal*, el cual intenta modelar las resonancias del tracto vocal. Luego la señal pasa por el filtro llamado *Radiación Labios*, el cual intenta modelar la dinámica agregada por los labios. Finalmente, la salida de esta serie de filtros es la señal del habla que puede ser medida, por ejemplo, por un micrófono.

Aunque este modelo es simple y de baja complejidad, su linealidad nos permite utilizar herramientas como la transformada Zeta [20]. En la Fig. 2.4 se muestra el diagrama de bloques que describe al modelo lineal de la producción del habla donde $E(z)$ es la transformada Zeta del tren de impulsos que excita al conjunto de filtros en serie para dar como resultado la señal del habla representada por su transformada Zeta $S(z)$.

En la próxima sección nos concentraremos en estudiar algunos de los modelos más utilizados para representar a $G(z)$. Por su parte, los filtros $V(z)$ y $R(z)$ no serán necesarios

²En el caso de los sonidos sin voz, la excitación es una señal de ruido aleatorio con un espectro plano.

de estudiar pues no son utilizados en el presente trabajo.

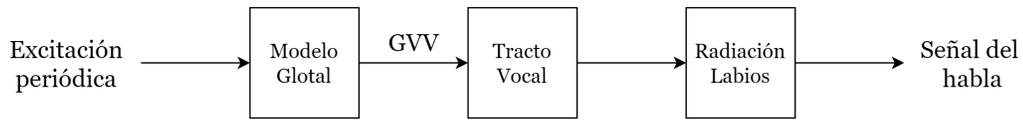


Figura 2.3: Modelo lineal de la producción del habla.

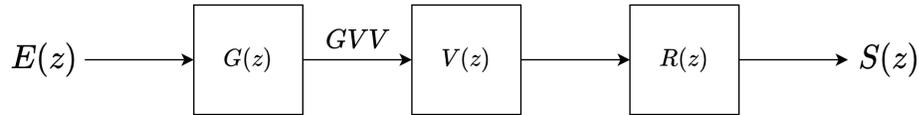


Figura 2.4: Modelo lineal de la producción del habla (Dominio z).

2.2. Modelo glotal

Para aplicar el enfoque propuesto en la presente tesis, es necesario un modelo que describa el flujo de aire glotal, esto es, necesitamos conocer $G(z)$. Este filtro tiene como objetivo darle la *forma de onda* de pulso glotal o flujo de aire glotal a un tren de impulsos. Así, dicho filtro, denominado *Modelo Glotal*, se ha modelado matemáticamente tanto en el dominio de la frecuencia como en el dominio del tiempo. En el primer caso se tiene una función de transferencia $G(z)$, mientras que en el segundo se tiene la respuesta a impulso g_t , las cuales están relacionadas por $G(z) = \mathcal{Z}\{g_t\}$. Es importante mencionar que en la literatura se encuentran descripciones del *Modelo Glotal* en tiempo continuo, esto es, en realidad se describen $G(s)$ o $g(t)$, los cuales tienen como equivalentes discretos a $G(z)$ y g_t , respectivamente.

Una de las primeras descripciones de $G(s)$ utilizadas en la *teoría fuente-filtro* de la producción del habla hace uso del efecto de la glotis en el dominio espectral para describir a $G(s)$ como un filtro pasabajo con polos en el eje real negativo [28]. Debido a lo anterior, en este trabajo se le denominará *Modelo de Butterworth* a dicha descripción de $G(s)$. Por otro lado, en [29] se intenta modelar la forma de onda del flujo de aire glotal en el dominio del tiempo, utilizando diversas descripciones matemáticas de $g(t)$ que se asemejen a las formas de onda observadas experimentalmente. El autor analiza 6 expresiones distintas para describir matemáticamente la forma de onda del flujo de aire glotal (lo que es equivalente a describir la respuesta a impulso $g(t)$), las cuales son referidas como A, B, C, D, E y F. Posteriormente, en [28], se realiza una revisión más actualizada de diversos modelos que sirven para modelar el espectro del flujo de aire glotal, estudiándolos tanto en el dominio del tiempo como de la frecuencia, donde se analizan las 4 expresiones más comunes referidas como *KLGLOTT88*, *Rosenberg C³*, *R++* y *LF*.

A continuación, se presentarán los modelos de Butterworth y Rosenberg C.

³Este es el modelo C de [29].

2.2.1. Modelo de Butterworth

La teoría fuente-filtro considera el modelo glotal en que el flujo de aire glotal es la salida de un filtro pasabajo excitado por un tren de impulsos. Fant [27] utilizó un filtro con cuatro polos en el eje real negativo:

$$G(s) = \frac{U_0}{\prod_{i=1}^4 (1 - s/s_{ri})} \quad (2.1)$$

donde $|s_{r1}| \simeq |s_{r2}| = 2\pi 100$ rad/s, $|s_{r3}| = 2\pi 2000$ rad/s, $|s_{r4}| = 2\pi 4000$ rad/s y U_0 es un factor de ganancia. De acuerdo a Fant, s_{r3} y s_{r4} son fijos, mientras que s_{r1} y s_{r2} se utilizan para considerar la variabilidad con respecto al hablante. Además, se incluye la frecuencia fundamental f_0 , dando un total de 6 parámetros [28]. Sin embargo, debido a que la linealidad del modelo acústico sólo se cumple para frecuencias menores a 4000 Hz, el modelo es simplificado para sólo considerar los polos s_{r1} y s_{r2} , resultando en un modelo de 3 parámetros: U_0 , f_0 y $s_{r1} \simeq s_{r2}$. Por lo anterior, las características espectrales del pulso glotal son equivalentes a la respuesta en frecuencia de un filtro de segundo orden, el cual puede ser descrito por un filtro de Butterworth cuya función de transferencia normalizada por la frecuencia de corte ω_c está dada por [30]:

$$H(\bar{s}) = \frac{H_0}{\bar{s}^2 + \sqrt{2}\bar{s} + 1} = \frac{H_0}{(\bar{s} - \bar{s}_0)(\bar{s} - \bar{s}_1)} \quad (2.2)$$

donde H_0 es la ganancia a continua y los polos están dados por $\bar{s}_0 = (-1 + j)/\sqrt{2}$ y $\bar{s}_1 = -(1 + j)/\sqrt{2}$. Si consideramos $\bar{s} = s/\omega_c$ obtenemos la función de transferencia del filtro de Butterworth no normalizado:

$$H(s) = \frac{H_0 \omega_c^2}{(s - s_0)(s - s_1)} \quad (2.3)$$

con $s_0 = \omega_c \bar{s}_0$ y $s_1 = \omega_c \bar{s}_1$. Posteriormente, es posible *muestrear* esta función de transferencia para obtener su equivalente discreto. Lo anterior se realiza mediante el mapeo de los polos continuos s_0 y s_1 a los polos discretos z_0 y z_1 como [31]:

$$z_1 = e^{s_1 T} \quad \text{y} \quad z_2 = e^{s_2 T} \quad (2.4)$$

donde T es el período de muestreo. Así, tenemos que la función de transferencia en tiempo discreto está dada por

$$G(z) = \frac{G_0}{(z - z_1)(z - z_2)} = \frac{H_0 z^{-2}}{1 - (z_1 + z_2)z^{-1} + (z_1 z_2)z^{-2}} \quad (2.5)$$

con G_0 tal que $G(1) = H(0)$ para que se conserve la ganancia a continua. La respuesta a impulso de este sistema se muestra en la Fig. 2.5 con $T = 5 \times 10^{-5}$ s y $\omega_c = 2\pi 200$ rad/s.

2.2.2. Modelo Rosenberg C

A partir de mediciones experimentales, Rosenberg [29] estudia varias descripciones matemáticas del flujo de aire glotal en el dominio del tiempo. Entre varias de las descripciones realizadas (6 en total), una que recibió mayor atención [28] es referida en dicho artículo con la letra C, por lo que en este trabajo de tesis dicho modelo será denominado *Rosenberg C*. Este modelo contiene 2 parámetros, los cuales se describen en la Tabla 2.1.

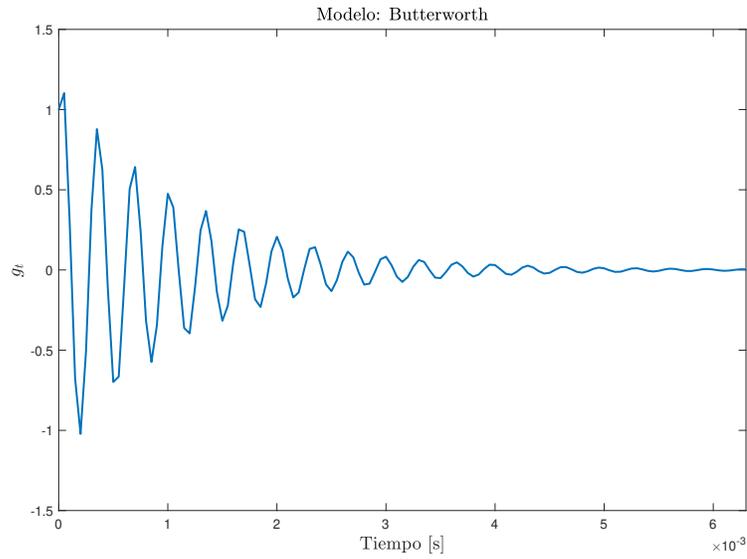


Figura 2.5: Modelo de Butterworth. Respuesta a impulso g_t obtenida del modelo $G(z)$ con período de muestreo igual a 20 kHz.

Parámetros Rosenberg C	
Parámetro	Descripción
T_1	Tiempo de fase de apertura.
T_2	Tiempo de fase de cierre.

Tabla 2.1: Parámetros utilizados por el modelo de Rosenberg C.

En este modelo la forma de onda del flujo de aire glotal g_t , está dada en tiempo discreto por [26, 28]:

$$g_t = \begin{cases} \frac{1}{2} \left[1 - \cos \left(\frac{\pi t}{N_1} \right) \right] & 0 \leq t \leq N_1 - 1 \\ \cos \left(\frac{\pi(t-N_1)}{2N_2} \right) & N_1 \leq t \leq N_1 + N_2 - 2 \\ 0 & \text{en otro caso} \end{cases} \quad (2.6)$$

donde⁴ $N_1 = \lceil T_1/T \rceil$, $N_2 = \lceil T_2/T \rceil$ y T es el período de muestreo. La Fig. 2.6 muestra un ejemplo de g_t usando como datos $T = 5 \times 10^{-5}$ s, $T_1 = 0.0011$ s (o $N_1 = 22$) y $T_2 = 0.00045$ s (o $N_2 = 9$).

2.3. Características aerodinámicas del flujo de aire glotal

Las características aerodinámicas extraídas del flujo de aire glotal son de vital importancia para la evaluación de la función vocal. Éstas han mostrado ser una buena

⁴Aquí $\lceil x \rceil$ es la función techo, la cual entrega el mínimo número entero igual o mayor a x .

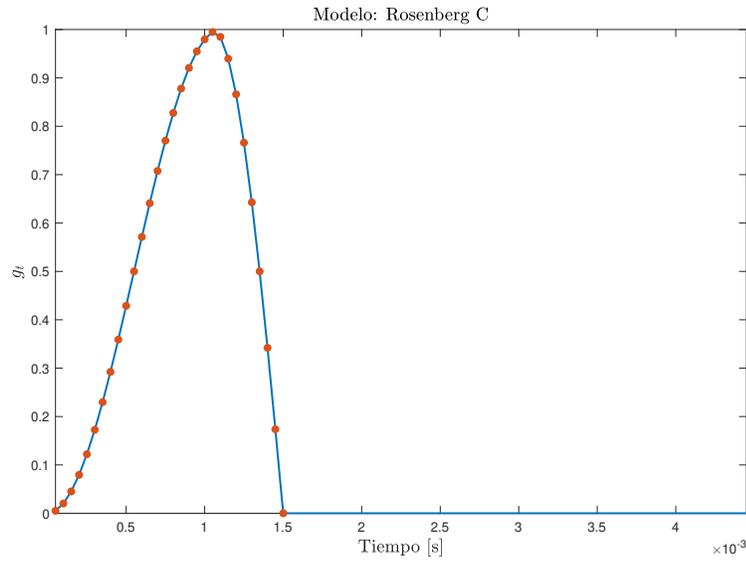


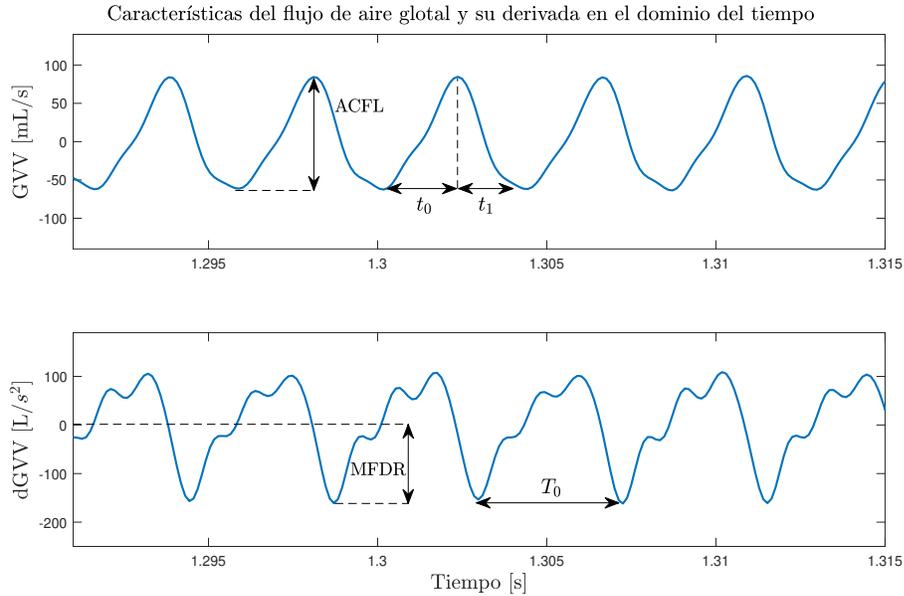
Figura 2.6: Modelo Rosenberg C.

medida objetiva para el correcto diagnóstico de la hiperfunción vocal de tipo fonotraumática (PVH) y de tipo no fonotraumática (NPVH) [7, 8]. En general, pacientes con PVH han mostrado valores anormalmente elevados de la *amplitud peak-to-peak del flujo no estacionario* (AC Flow o ACFL) y de la *razón máxima de declinación de flujo* (MFDR), lo cual ha sido interpretado como el reflejo de un mayor potencial de traumatismo en el tejido de las cuerdas vocales, lo que contribuye a la presencia crónica de lesiones de las cuerdas vocales y la disfonía asociada [32]. Los pacientes con PVH también tienden a tener valores elevados del *cociente de apertura* (OQ) lo que ha sido atribuido a la obstrucción del cierre glotal debido a la presencia de lesiones en las cuerdas vocales. Por otro lado, pacientes con NPVH también han mostrado valores anormalmente elevados de OQ, aunque sin un acompañamiento en el incremento de ACFL y MFDR, lo que fue asociado con fonación ineficaz y disfonía, aunque con un menor potencial de causar trauma al tejido de las cuerdas vocales [32].

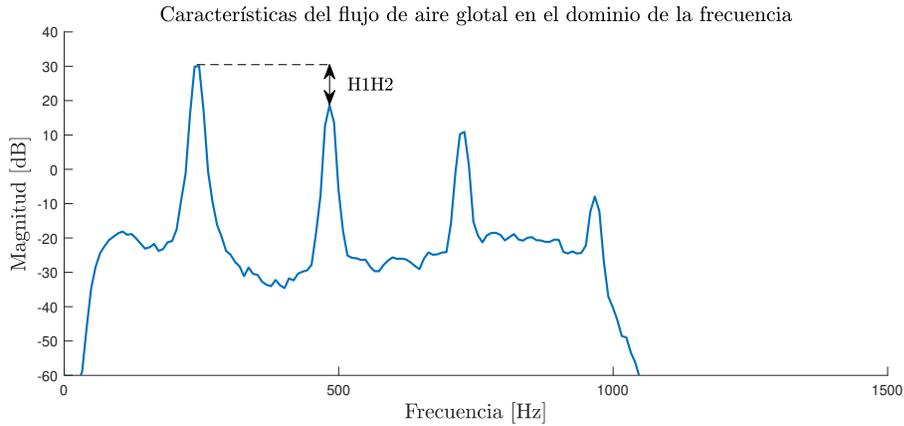
Para definir las distintas características aerodinámicas extraídas del flujo de aire glotal y su derivada, sea $\mathbf{u}^g = \{u_0^g, u_1^g, \dots, u_{n_0-1}^g\}$ el vector que contiene muestras del flujo de aire glotal en un solo ciclo glotal, donde n_0 es el número de muestras en un ciclo. Además, considérese la Fig. 2.7 como referencia para comprender las siguientes medidas del flujo de aire glotal y su derivada [32]:

1. **Frecuencia fundamental (f_0):** Primer armónico de la señal de flujo de aire periódica. Puede ser formulada como $f_0 = f_s/n_0$, donde f_s es la frecuencia de muestreo. Su unidad es el Hz.
2. **Cociente de rapidez (SQ):** Proporción (razón) entre el tiempo de apertura y tiempo de cierre del pulso glotal.

$$SQ = \frac{t_1}{t_2} \quad (\%) \quad (2.7)$$



(a) Medidas en el dominio del tiempo.



(b) Medidas en el dominio de la frecuencia.

Figura 2.7: Características aerodinámicas del flujo de aire glotal estimado y su derivada. Se indica el valor *peak-to-peak* de la amplitud (ACFL), el tiempo de apertura (t_1), tiempo de cierre (t_2), razón máxima de declinación de flujo (MFDR), el período de un ciclo glotal calculado entre *peaks* consecutivos de la derivada del flujo de aire glotal estimado (T_0) y la razón entre el primer y segundo armónico (H1H2).

3. **Cociente de apertura (OQ):** Proporción de tiempo en el cual las cuerdas vocales están en fase abierta con respecto a un período ($T_0 = 1/f_0 = n_0/f_s$).

$$\text{OQ} = \frac{t_1 + t_2}{T_0} = (t_1 + t_2) \cdot f_0 \quad (\%) \quad (2.8)$$

4. **Amplitud peak-to-peak del flujo de aire glotal (ACFL):**

$$\text{ACFL} = \text{máx } \mathbf{u}^g - \text{mín } \mathbf{u}^g \quad (\text{mL/s}) \quad (2.9)$$

5. **Razón máxima de declinación de flujo (MFDR):** Magnitud del *peak* negativo de la derivada temporal del flujo de aire glotal.

$$\text{MFDR} = |\text{mín } \Delta \mathbf{u}^g| \quad (\text{L/s}^2) \quad (2.10)$$

6. **Razón entre el primer y segundo armónico (H1H2):** Es la diferencia (en dB) entre las magnitudes del primer y segundo armónico. Es estimada a partir de la magnitud del espectro de u_t^g , $U_g(f)$, el cual es calculado mediante la Transformada Discreta de Fourier (DFT).

$$\text{H1H2} = 20 \log_{10} \left(\frac{|U_g(f_0)|}{|U_g(2f_0)|} \right) \quad (\text{dB}) \quad (2.11)$$

7. **Cociente de amplitud normalizado (NAQ):** Es una medida para parametrizar la fase de cierre glotal mediante el uso de ACFL y MFDR, normalizado mediante un período. NAQ es una estimación robusta para la fase de cierre (CQ).

$$\text{NAQ} = \frac{\text{ACFL}}{\text{MFDR}} \cdot \frac{1}{T_0} = \frac{\text{ACFL}}{\text{MFDR}} \cdot f_0 \propto \text{CQ} = \frac{t_2}{T_0} \quad (-) \quad (2.12)$$

8. **Factor de riqueza armónica (HRF):** Proporción entre la magnitud del primer armónico con respecto a la suma de las magnitudes de los demás armónicos.

$$\text{HRF} = 20 \log_{10} \left(\frac{\sum_{n>1} |U_g(nf_0)|}{|U_g(f_0)|} \right) \quad (\text{dB}) \quad (2.13)$$

2.4. Identificación de sistemas

La *Identificación de Sistemas* trata el problema de construir modelos matemáticos de sistemas dinámicos mediante el uso de datos experimentales. Estos datos, en general, corresponden a mediciones de las señales de entrada y salida del sistema, donde estas señales pueden pertenecer al dominio del tiempo o al dominio de la frecuencia [33, 34]. Si consideramos el esquema de la Fig. 2.8, podemos entender la identificación de sistemas como aquella área que trata de describir, matemáticamente, el bloque denominado *Sistema*, haciendo uso de datos experimentales de la entrada \mathbf{u}_t y de la salida \mathbf{y}_t y teniendo en cuenta la presencia del ruido no medible \mathbf{e}_t .

Un modelo matemático básico y ampliamente utilizado para describir sistemas son las ecuaciones diferenciales, en el caso continuo, y las ecuaciones de diferencias, en el caso discreto. Cuando estas ecuaciones son lineales, herramientas poderosas como la transformada de Laplace (o el operador de Heaviside ρ [20]) y la transformada Zeta (o el operador adelanto

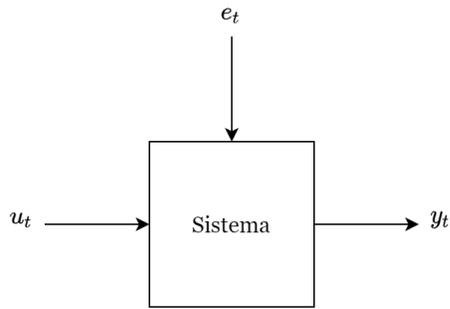


Figura 2.8: Sistema general con entrada, salida y ruido.

q [20]) sirven para analizarlas desde una perspectiva más útil e intuitiva al momento de obtener características del sistema que describen, por ejemplo, la estabilidad, la velocidad, entre otras [20]. En específico, es posible definir *funciones de transferencia*, las cuales describen matemáticamente la relación entrada-salida de los sistemas en los dominios s (Laplace) y z (Zeta). Así, un sistema discreto, lineal e invariante en el tiempo (LTI), puede ser descrito mediante el uso de funciones de transferencia, tal como lo muestra el diagrama de bloques de la Fig. 2.9 o, alternativamente, la siguiente ecuación

$$y_t = G(q)u_t + H(q)e_t \quad (2.14)$$

donde

$$G(q) = \sum_{k=0}^{\infty} g_k q^{-k} \quad (2.15)$$

$$H(q) = \sum_{k=0}^{\infty} h_k q^{-k} \quad (2.16)$$

con $\{g_k\}$ y $\{h_k\}$ siendo las respuestas a impulsos de cada sub-sistema.

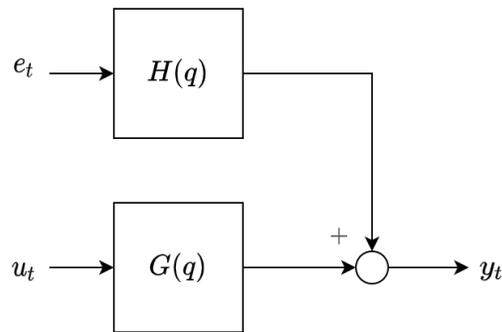


Figura 2.9: Diagrama de bloques usando funciones de transferencia.

Los modelos obtenidos mediante el uso de funciones de transferencia son llamados *modelos de entrada-salida* pues, como se puede observar de la ec. (2.14), relacionan

directamente la entrada u_t con la salida y_t , sin variables intermedias. Una forma alternativa de representar matemáticamente un sistema es mediante los modelos en *variables de estado*, los cuales describen las relaciones de todas las variables pertenecientes al sistema, entre las que se incluyen entradas, salidas y variables internas (denominadas *estados del sistema*). Un modelo en variables de estado de un sistema discreto, lineal, invariante en el tiempo y determinístico-estocástico se puede describir mediante las siguientes ecuaciones

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \quad (2.17)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t \quad (2.18)$$

donde $\mathbf{x}_t \in \mathbb{R}^n$ es el vector de estados del sistema, $\mathbf{u}_t \in \mathbb{R}^p$ es el vector de entradas del sistema, $\mathbf{y}_t \in \mathbb{R}^m$ es el vector de salidas del sistema, $\mathbf{w}_t \in \mathbb{R}^n$ es el vector de ruido de proceso y $\mathbf{v}_t \in \mathbb{R}^m$ es el vector de ruidos de salida, mientras que $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, $\mathbf{C} \in \mathbb{R}^{m \times n}$ y $\mathbf{D} \in \mathbb{R}^{m \times p}$ son las matrices de estado, de entrada, de salida y de paso directo, respectivamente. Aquí, los ruidos \mathbf{w}_t y \mathbf{v}_t los supondremos Gaussianos de media 0 con matrices de covarianza \mathbf{Q} y \mathbf{R} , respectivamente, y matriz de covarianza cruzada \mathbf{S} .

Una manera alternativa de representar el modelo en variables de estado anterior, es mediante su forma en innovaciones, donde los ruidos \mathbf{w}_t y \mathbf{v}_t son *equivalentes* a un ruido \mathbf{e}_t . Así, la forma en innovaciones de (2.17)-(2.18) está dada por

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{K}\mathbf{e}_t \quad (2.19)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{e}_t. \quad (2.20)$$

Los modelos descritos con funciones de transferencia y descritos en variables de estado son formas equivalentes de representar matemáticamente un sistema, con la única diferencia de que los últimos incluyen mayor información sobre las dinámicas de éste (debido a las variables de estado). En efecto, la relación entre las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} y las funciones de transferencia $G(q)$ y $H(q)$ está dada por

$$G(q) = \mathbf{C}(q\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \quad (2.21)$$

$$H(q) = \mathbf{C}(q\mathbf{I} - \mathbf{A})^{-1}\mathbf{K} + \mathbf{I}. \quad (2.22)$$

Por lo anterior, el problema de la identificación de sistemas se puede parafrasear como *“encontrar las funciones de transferencia $G(q)$ y $H(q)$ o, alternativamente, las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} , utilizando mediciones en el dominio del tiempo, representadas por u_t y y_t o en el dominio de la frecuencia, representadas por $U(e^{j\omega}) = \mathcal{F}\{u_t\}$ y $Y(e^{j\omega}) = \mathcal{F}\{y_t\}$, que mejor se ajusten a las mediciones obtenidas”*. En este punto, se debe mencionar que existen dos métodos predominantes para la búsqueda de dichas incógnitas: los *métodos paramétricos* y los *métodos no-paramétricos*. Los primeros, como el nombre lo indica, suponen que $G(q)$ y $H(q)$ o, alternativamente, \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} tienen cierta estructura y que dependen de un vector de parámetros $\boldsymbol{\theta}$ que los definen. Esto es, $G(q)$ y $H(q)$ se pueden escribir como $G(q, \boldsymbol{\theta})$ y $H(q, \boldsymbol{\theta})$ y \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} como $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{C}(\boldsymbol{\theta})$, $\mathbf{D}(\boldsymbol{\theta})$ y $\mathbf{K}(\boldsymbol{\theta})$. Por otro lado, los métodos no-paramétricos buscan obtener $G(q)$ y $H(q)$ o, alternativamente, \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} sin el uso explícito del vector de parámetros $\boldsymbol{\theta}$ [33].

Los métodos paramétricos dependen de la parametrización de las funciones de transferencia o de las matrices de estado, es decir, de la forma en que éstas incógnitas dependen de los parámetros. En el caso de las funciones de transferencia, la parametrización se realiza de manera natural mediante el uso de funciones racionales en los parámetros, tal

como se ve en las ec. (2.23) y (2.24)

$$G(q^{-1}, \boldsymbol{\theta}) = \frac{B(q^{-1}, \boldsymbol{\theta})}{A(q^{-1}, \boldsymbol{\theta})} = \frac{1 + b_1 q^{-1} + \dots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \dots + a_{na} q^{-na}} \quad (2.23)$$

$$H(q^{-1}, \boldsymbol{\theta}) = \frac{C(q^{-1}, \boldsymbol{\theta})}{D(q^{-1}, \boldsymbol{\theta})} = \frac{1 + c_1 q^{-1} + \dots + c_{nc} q^{-nc}}{1 + d_1 q^{-1} + \dots + d_{nd} q^{-nd}} \quad (2.24)$$

con el vector de parámetros $\boldsymbol{\theta}$ dado por

$$\boldsymbol{\theta} = [a_1, \dots, a_{na}, b_1, \dots, b_{nb}, c_1, \dots, c_{nc}, d_1, \dots, d_{nd}]^\top. \quad (2.25)$$

Sin embargo, en el caso de los modelos en variables de estado, parametrizar $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{C}(\boldsymbol{\theta})$, $\mathbf{D}(\boldsymbol{\theta})$ y $\mathbf{K}(\boldsymbol{\theta})$ no es trivial, pues el trabajar con matrices implica trabajar con todas las componentes de éstas. Así, si parametrizamos completamente cada una de estas matrices, tendríamos un total de $n^2 + np + 2mn + mp$ parámetros a determinar, lo que en general dará una sobre-parametrización. A pesar de que la literatura ha estudiado parametrizaciones (ver [33]) que cubren todos los sistemas posibles usando la cantidad precisa de parámetros (aunque con posiciones fijas), en [35] se defiende la idea de la parametrización completa debido a que con ella se cubren todos los casos posibles de sistemas, sin fijar la estructura de las matrices.

Para entender los métodos utilizados en la presente tesis, en la siguientes subsecciones se describen brevemente los métodos existentes en la literatura para la obtención de \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} y \mathbf{K} a partir de datos experimentales. Por simplicidad, sólo se considerarán los métodos que utilizan datos en el dominio del tiempo y, en algunos casos, sólo se considerará un sistema determinístico, esto es, con $\mathbf{w}_t = \mathbf{v}_t = 0$ (o, alternativamente, $\mathbf{e}_t = 0$).

2.4.1. Teoría de realizaciones

Si consideramos la ec. (2.15) podemos notar que $G(q)$ está dado por una sumatoria infinita de la respuesta a impulso $\{g_k\}$. Al mismo tiempo, en la ec. (2.21) podemos ver que $G(q)$ está descrito mediante el uso de matrices de orden finito. Por lo anterior, cabría preguntarnos cómo es posible obtener una descripción finita a partir de una descripción infinita. En un intento de responder la pregunta anterior es que nace la *Teoría de realizaciones*, la cual trata el problema de obtener las matrices de orden finito \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{D} a partir de la respuesta a impulso infinita $\{g_k\}$ sin la necesidad de parametrizar las matrices en cuestión. La primera solución a este problema, para el caso determinístico, fue entregada por Ho y Kalman en [36], mientras que Akaike, en [37], entregó una solución para el caso estocástico. Posteriores mejoras, mediante el uso de la *descomposición en valores singulares (SVD)*, fueron realizadas por Zeiger y McEwen en [38].

Para comprender los métodos utilizados en la teoría de realizaciones [35, 36, 38], primero debemos entender la relación entre las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{D} y la respuesta a impulso $\{g_k\}$, donde consideraremos un sistema MIMO (*Multiple-Input Multiple-Output*) con m salidas y p entradas (con lo que $\mathbf{g}_k \in \mathbb{R}^{m \times p}$):

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \quad (2.26)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \quad (2.27)$$

donde supondremos que el sistema es observable y controlable y, por lo tanto, es de orden mínimo n .

Luego, si $\mathbf{u}_t = \delta_t$ (delta de Kronecker) y $\mathbf{x}_0 = 0$, tenemos que $\mathbf{y}_t = \mathbf{g}_t$ y, al iterar sobre las ecuaciones anteriores, obtenemos

$$\mathbf{g}_t = \begin{cases} \mathbf{D}, & t = 0 \\ \mathbf{C}\mathbf{A}^{t-1}\mathbf{B}, & t > 0 \end{cases} \quad (2.28)$$

donde podemos ver que $\mathbf{D} = \mathbf{g}_0$, por lo que de aquí en adelante nos concentraremos en encontrar las matrices \mathbf{A} , \mathbf{B} y \mathbf{C} .

Ahora, tres matrices importantes en la teoría de realizaciones son la *matriz de Hankel* $\mathcal{H}_{ij} \in \mathbb{R}^{pi \times mj}$, la cual es construida a partir de los coeficientes de la respuesta a impulso, y las *matrices extendidas de observabilidad* y *controlabilidad* $\mathcal{O}_i \in \mathbb{R}^{pi \times n}$ y $\mathcal{C}_j \in \mathbb{R}^{n \times mj}$. Estas tres matrices están dadas por

$$\mathcal{H}_{ij} = \begin{pmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_j \\ \mathbf{g}_2 & \mathbf{g}_3 & \cdots & \mathbf{g}_{j+1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{g}_i & \mathbf{g}_{i+1} & \cdots & \mathbf{g}_{i+j-1} \end{pmatrix} \quad (2.29)$$

$$\mathcal{O}_i = \begin{pmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{i-1} \end{pmatrix} \quad (2.30)$$

$$\mathcal{C}_j = \begin{pmatrix} \mathbf{B} & \mathbf{A}\mathbf{B} & \cdots & \mathbf{A}^{j-1}\mathbf{B} \end{pmatrix} \quad (2.31)$$

donde, si consideramos (2.28), es posible ver que las tres matrices anteriores cumplen con la siguiente igualdad

$$\mathcal{H}_{ij} = \mathcal{O}_i \mathcal{C}_j. \quad (2.32)$$

A continuación se describe el algoritmo descrito por Zeiger y McEwen en [38], el cual sigue las ideas algebraicas de [36] y es utilizado para obtener las matrices \mathbf{A} , \mathbf{B} y \mathbf{C} mediante la descomposición por valores singulares y el uso de la respuesta a impulso $\{\mathbf{g}_k\}$.

Así, sea $\underline{\mathcal{H}}_{ij} \in \mathbb{R}^{pi \times mj}$ la matriz de Hankel con \mathbf{g}_2 como elemento superior izquierdo pero con el mismo tamaño que \mathcal{H}_{ij} y supongamos que $\min(i, j) \geq n$, donde n es el orden del sistema, para garantizar que \mathcal{O}_i y \mathcal{C}_j sean de rango completo. Luego, definamos la SVD

$$\underline{\mathcal{H}}_{ij} = \begin{pmatrix} \mathbf{U}_s & \mathbf{U}_o \end{pmatrix} \begin{pmatrix} \Sigma_s & \mathbf{0} \\ \mathbf{0} & \Sigma_o \end{pmatrix} \begin{pmatrix} \mathbf{V}_s^\top \\ \mathbf{V}_o^\top \end{pmatrix} \quad (2.33)$$

donde Σ_s es la matriz diagonal que contiene los n valores singulares de mayor valor. La SVD descompone la matriz de Hankel en dos términos, el término $\mathbf{U}_s \Sigma_s \mathbf{V}_s^\top$ y el término $\mathbf{U}_o \Sigma_o \mathbf{V}_o^\top$, donde si la respuesta a impulso $\{\mathbf{g}_k\}$ no contiene ruido, Σ_o es igual a cero y, por tanto, el último término se hace cero, lo que significa que es posible obtener el orden del

sistema considerando los n valores singulares distintos a 0. Luego, se utiliza la SVD para factorizar la matriz de Hankel como

$$\mathcal{H}_{ij} = (\mathbf{U}_s \boldsymbol{\Sigma}_s^{1/2})(\boldsymbol{\Sigma}_s^{1/2} \mathbf{V}_s^\top) \quad (2.34)$$

donde podemos tomar

$$\hat{\mathcal{O}}_i = \mathbf{U}_s \boldsymbol{\Sigma}_s^{1/2} \quad \text{y} \quad \hat{\mathcal{C}}_j = \boldsymbol{\Sigma}_s^{1/2} \mathbf{V}_s^\top \quad (2.35)$$

como estimaciones de las matrices de observabilidad y controlabilidad, respectivamente. Se puede mostrar que existe una matriz no-singular $\mathbf{T} \in \mathbb{R}^{n \times n}$ tal que

$$\mathcal{O}_i = \hat{\mathcal{O}}_i \mathbf{T} \quad \text{y} \quad \mathcal{C}_j = \mathbf{T}^{-1} \hat{\mathcal{C}}_j \quad (2.36)$$

donde \mathcal{O}_i y \mathcal{C}_j son las matrices extendidas de observabilidad y controlabilidad construidas a partir de la realización (2.26)-(2.27). La factorización obtenida con la SVD entonces entrega las matrices de observabilidad y controlabilidad para alguna realización $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{D}})$ del sistema (2.26)-(2.27). Luego, notamos que la matriz \mathcal{H}_{ij} puede escribirse como

$$\mathcal{H}_{ij} = \mathcal{O}_i \mathbf{A} \mathcal{C}_j = \hat{\mathcal{O}}_i \hat{\mathbf{A}} \hat{\mathcal{C}}_j \quad (2.37)$$

Así, una estimación de $\hat{\mathbf{A}}$ puede ser obtenida como

$$\hat{\mathbf{A}} = \hat{\mathcal{O}}_i^\dagger \mathcal{H}_{ij} \hat{\mathcal{C}}_j^\dagger \quad (2.38)$$

donde $(\cdot)^\dagger$ denota la pseudo-inversa de Moore-Penrose. Si utilizamos la SVD original de \mathcal{H}_{ij} , ec. (2.34) y (2.35), podemos ver que la ecuación anterior puede ser escrita como

$$\hat{\mathbf{A}} = \boldsymbol{\Sigma}_s^{-1/2} \mathbf{U}_s^\top \mathcal{H}_{ij} \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1/2}. \quad (2.39)$$

Finalmente, \mathbf{B} y \mathbf{C} pueden ser obtenidos a partir de las matrices estimadas de observabilidad y controlabilidad como

$$\hat{\mathbf{B}} = \hat{\mathcal{C}}_j \left(\mathbf{I}_m \quad \mathbf{0}_{m \times (j-1)m} \right)^\top \quad (2.40)$$

$$\hat{\mathbf{C}} = \left(\mathbf{I}_p \quad \mathbf{0}_{p \times (i-1)p} \right) \hat{\mathcal{O}}_i. \quad (2.41)$$

Este algoritmo ha sido implementado fácilmente en el software MATLAB.

2.4.2. Métodos de subespacios

Como se vio anteriormente, la teoría de realizaciones busca resolver el problema de obtener las matrices de la representación en variables de estado a partir de la respuesta a impulso de un sistema. Sin embargo, no siempre es posible medir la respuesta a impulso de un sistema, por lo que se vuelve necesario extender la solución del problema para el caso donde se tienen disponibles datos de entrada y salida de señales generales. Por lo anterior, es que con base a la teoría de la realización surgen los métodos de *Identificación de Sistemas en Variables de Estado basados en Subespacios* o *4SID* (por sus siglas en inglés) [39–41]. Estos métodos de subespacios hacen uso de ideas similares a las utilizadas en la teoría de realizaciones, tales como proyecciones y SVD, por ejemplo. Es por esta *herencia*, que

un primer enfoque que surgió en los métodos de subespacios, denominado *4SID basados en realizaciones*, fue el de obtener una estimación de la respuesta a impulso a partir de los datos de entrada y salida, para luego utilizar directamente métodos de la teoría de realizaciones [41].

El otro enfoque, que es al que nos referiremos en la presente tesis, fue denominado *4SID directo* [41] pues intenta obtener las matrices del sistema sin estimar la respuesta a impulso. Así, la idea básica de los métodos 4SID directos, es como se describe a continuación. Consideremos el siguiente modelo en variables de estado general

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \quad (2.42)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t \quad (2.43)$$

el cual, se puede escribir alternativamente como

$$\bar{\mathbf{y}}_t = \Theta\varphi_t + \boldsymbol{\eta}_t \quad (2.44)$$

con

$$\bar{\mathbf{y}}_t = \begin{pmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_t \end{pmatrix} \quad \Theta = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \quad \varphi_t = \begin{pmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{pmatrix} \quad \boldsymbol{\eta}_t = \begin{pmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{pmatrix}. \quad (2.45)$$

Se tienen como datos \mathbf{u}_t y \mathbf{y}_t para $t = 0, 1, \dots, N-1$. Así, si suponemos por un momento que conocemos \mathbf{x}_t para $t = 0, 1, \dots, N$, entonces el problema anterior se convierte en un problema de mínimos cuadrados, cuya solución está dada por:

$$\Theta = \left(\sum_{t=0}^{N-1} \bar{\mathbf{y}}_t \varphi_t^\top \right) \left(\sum_{t=0}^{N-1} \varphi_t \varphi_t^\top \right)^{-1}. \quad (2.46)$$

Por lo anterior, los métodos de subespacios directos se enfocan en la obtención de una estimación $\bar{\mathbf{x}}_t$ del estado \mathbf{x}_t a partir de los datos disponibles de entrada y salida. A continuación describiremos cómo obtener una estimación del estado \mathbf{x}_t en el caso de un sistema determinístico, esto es, con $\mathbf{w}_t = \mathbf{v}_t = 0$.

Para comenzar, definimos las matrices de entrada de Hankel $\mathbf{U}_{0:2i-1} \in \mathbb{R}^{2mi \times j}$, $\mathbf{U}_p \in \mathbb{R}^{mi \times j}$ y $\mathbf{U}_f \in \mathbb{R}^{mi \times j}$ como

$$\mathbf{U}_{0:2i-1} = \begin{pmatrix} \mathbf{u}_0 & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{j-1} \\ \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \cdots & \mathbf{u}_j \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_{i-1} & \mathbf{u}_i & \mathbf{u}_{i+1} & \cdots & \mathbf{u}_{i+j-2} \\ \hline \mathbf{u}_i & \mathbf{u}_{i+1} & \mathbf{u}_{i+2} & \cdots & \mathbf{u}_{i+j-1} \\ \mathbf{u}_{i+1} & \mathbf{u}_{i+2} & \mathbf{u}_{i+3} & \cdots & \mathbf{u}_{i+j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_{2i-1} & \mathbf{u}_{2i} & \mathbf{u}_{2i+1} & \cdots & \mathbf{u}_{2i+j-2} \end{pmatrix} \quad (2.47)$$

$$= \begin{pmatrix} \mathbf{U}_{0:i-1} \\ \mathbf{U}_{i:2i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_p \\ \mathbf{U}_f \end{pmatrix} \quad (2.48)$$

donde los subíndices p y f se refieren a *pasado* y *futuro*, respectivamente, el número de filas i es un índice definido por el usuario, el cual debe ser al menos igual al orden n del sistema que uno quiere identificar y el número de columnas j típicamente es igual a $N - 2i + 1$, lo que implica que todos los datos disponibles son utilizados. De forma similar, definimos las matrices de salida de Hankel $\mathbf{Y}_{0:2i-1} \in \mathbb{R}^{2pi \times j}$, $\mathbf{Y}_p \in \mathbb{R}^{pi \times j}$ y $\mathbf{Y}_f \in \mathbb{R}^{pi \times j}$.

Además, definimos las matrices de Hankel que consisten de entradas y salidas como

$$\mathbf{W}_{0:i-1} = \begin{pmatrix} \mathbf{U}_{0:i-1} \\ \mathbf{Y}_{0:i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_p \\ \mathbf{Y}_p \end{pmatrix} = \mathbf{W}_p \quad (2.49)$$

mientras que la secuencia de estados $\mathbf{X}_i \in \mathbb{R}^{n \times j}$ se define como

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_i & \mathbf{x}_{i+1} & \cdots & \mathbf{x}_{i+j-2} & \mathbf{x}_{i+j-1} \end{pmatrix} \quad (2.50)$$

donde el subíndice i denota el subíndice del primer elemento de la secuencia de estados. Adicionalmente, definimos la secuencia de estados pasados \mathbf{X}_p y de estados futuros \mathbf{X}_f como

$$\mathbf{X}_p = \mathbf{X}_0 \quad \mathbf{X}_f = \mathbf{X}_i. \quad (2.51)$$

Por otro lado, la matriz extendida de observabilidad $\mathbf{O}_i \in \mathbb{R}^{pi \times n}$ es como en (2.30), mientras que la matriz extendida de controlabilidad invertida $\bar{\mathbf{C}}_i \in \mathbb{R}^{n \times mi}$ es como en (2.31) pero partiendo desde $\mathbf{A}^{i-1}\mathbf{B}$ hasta llegar a \mathbf{B} . Finalmente, la matriz triangular inferior de Toeplitz $\mathbf{H}_i \in \mathbb{R}^{pi \times mi}$ está definida como

$$\mathbf{H}_i = \begin{pmatrix} \mathbf{D} & 0 & 0 & \cdots & 0 \\ \mathbf{CB} & \mathbf{D} & 0 & \cdots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{CA}^{i-2}\mathbf{B} & \mathbf{CA}^{i-3}\mathbf{B} & \mathbf{CA}^{i-4}\mathbf{B} & \cdots & \mathbf{D} \end{pmatrix} \quad (2.52)$$

De las definiciones anteriores es posible deducir el siguiente teorema.

Teorema 2.1 (Ecuaciones matriciales de entrada-salida).

$$\mathbf{Y}_p = \mathbf{O}_i \mathbf{X}_p + \mathbf{H}_i \mathbf{U}_p \quad (2.53)$$

$$\mathbf{Y}_f = \mathbf{O}_i \mathbf{X}_f + \mathbf{H}_i \mathbf{U}_f \quad (2.54)$$

$$\mathbf{X}_f = \mathbf{A}^i \mathbf{X}_p + \bar{\mathbf{C}}_i \mathbf{U}_p \quad (2.55)$$

Demostración. La demostración sigue directamente de la iteración sobre las ecuaciones de estado (2.42)-(2.43) con $\mathbf{w}_t = \mathbf{v}_t = 0$ y las definiciones anteriores. \blacksquare

Antes de continuar, es necesario entender los conceptos de *proyección ortogonal* y *persistencia de excitación*, los cuales se describen a continuación [39].

Definición 2.1 (Proyecciones Ortogonales [39]). El operador que proyecta el espacio fila de una matriz en el espacio fila de la matriz $\mathbf{B} \in \mathbb{R}^{q \times j}$ se denota como $\mathbf{\Pi}_B$ y está dado por

$$\mathbf{\Pi}_B = \mathbf{B}^\top (\mathbf{B}\mathbf{B}^\top)^\dagger \mathbf{B}. \quad (2.56)$$

Así, la proyección del espacio fila de la matriz $\mathbf{A} \in \mathbb{R}^{p \times j}$ en el espacio fila de la matriz \mathbf{B} se denota como \mathbf{A}/\mathbf{B} y está dado por

$$\mathbf{A}/\mathbf{B} = \mathbf{A}/\mathbf{\Pi}_B = \mathbf{A}\mathbf{B}^\top (\mathbf{B}\mathbf{B}^\top)^\dagger \mathbf{B}. \quad (2.57)$$

Finalmente, $\mathbf{\Pi}_{B^\perp}$ es el operador geométrico que proyecta el espacio fila de una matriz en el complemento ortogonal del espacio fila de la matriz \mathbf{B} de la siguiente manera

$$\mathbf{A}/\mathbf{B}^\perp = \mathbf{A}\mathbf{\Pi}_{B^\perp} \quad (2.58)$$

donde

$$\mathbf{\Pi}_{B^\perp} = \mathbf{I}_j - \mathbf{\Pi}_B. \quad (2.59)$$

Definición 2.2 (Persistencia de excitación [39]). La secuencia de entrada $\mathbf{u}_t \in \mathbb{R}^p$ es persistentemente excitante de orden $2i$ si la matriz de covarianza de entrada

$$\mathbf{R}_u = \lim_{j \rightarrow \infty} \frac{1}{j} [\mathbf{U}_{0:2i-1} \mathbf{U}_{0:2i-1}^\top] \quad (2.60)$$

tiene rango completo, el cual es $2pi$.

Con lo anterior, es posible derivar el siguiente teorema, el cual nos indica cómo obtener estimaciones del estado \mathbf{x}_t , las cuales están contenidas en la secuencia de estados \mathbf{X}_f , a partir de las mediciones de \mathbf{u}_t y \mathbf{y}_t .

Teorema 2.2 (Identificación determinística). *Supongamos que*

1. La entrada \mathbf{u}_t es persistentemente excitante de orden $2i$.
2. La intersección del espacio fila de \mathbf{U}_f y el espacio fila de \mathbf{X}_p es vacía.
3. Las matrices de peso definidas por el usuario $\mathbf{W}_1 \in \mathbb{R}^{mi \times mi}$ y $\mathbf{W}_2 \in \mathbb{R}^{j \times j}$ son tales que \mathbf{W}_1 es de rango completo y \mathbf{W}_2 cumple con $\text{rank}(\mathbf{W}_p) = \text{rank}(\mathbf{W}_p \mathbf{W}_2)$, donde \mathbf{W}_p es la matriz de Hankel que contiene las entradas y salidas pasadas.

Y consideremos la matriz \mathbf{O}_i definida como

$$\mathbf{O}_i = (\mathbf{Y}_f / \mathbf{U}_f^\perp) (\mathbf{W}_p / \mathbf{U}_f^\perp)^\dagger \mathbf{W}_p \quad (2.61)$$

junto a la SVD

$$\mathbf{W}_1 \mathbf{O}_i \mathbf{W}_2 = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^\top \\ \mathbf{V}_2^\top \end{pmatrix} \quad (2.62)$$

Entonces tenemos que:

1. La matriz \mathcal{O}_i está dada por

$$\mathcal{O}_i = \mathcal{O}_i \mathbf{X}_f \quad (2.63)$$

2. El orden del sistema es igual al número de valores singulares de la ec. (2.62) diferentes de cero.

3. La matriz extendida de observabilidad es igual a

$$\mathcal{O}_i = \mathbf{W}_1^{-1} \mathbf{U}_1 \mathbf{S}_1^{1/2} \mathbf{T} \quad (2.64)$$

con $\mathbf{T} \in \mathbb{R}^{n \times n}$ siendo una matriz no-singular arbitraria.

4. La secuencia de estados \mathbf{X}_f está dada por

$$\mathbf{X}_f = \mathcal{O}_i^\dagger \mathcal{O}_i \quad (2.65)$$

Demostración. Ver [39]. ■

Del teorema anterior, es necesario realizar dos comentarios: el primero es relativo a la elección de \mathbf{T} en el punto 3, donde usualmente se elige $\mathbf{T} = \mathbf{I}_n$, mientras que el segundo es relativo a la elección de las matrices de peso \mathbf{W}_1 y \mathbf{W}_2 , cuya elección define, en el caso determinístico-estocástico, ciertos algoritmos populares. Por ejemplo, con $\mathbf{W}_1 = \mathbf{I}_{m_i}$ y $\mathbf{W}_2 = \mathbf{I}_j$ se tiene el algoritmo **N4SID** (*Numerical algorithms for Subspace State Space System IDentification*), mientras que con $\mathbf{W}_1 = \mathbf{I}_{m_i}$ y $\mathbf{W}_2 = \mathbf{\Pi}_{U_f^\perp}$ se tiene el algoritmo **MOESP** (*Multivariable Output-Error State sPace*) [39].

Gracias al punto 4 del teorema anterior, es posible obtener $\mathbf{X}_f = \mathbf{X}_i$ a partir de las mediciones de \mathbf{u}_t y \mathbf{y}_t (las cuales están incluidas en \mathbf{Y}_f , \mathbf{U}_f y \mathbf{W}_p). Luego, mediante un razonamiento similar al utilizado en la demostración del teorema anterior, es posible mostrar que

$$\mathcal{O}_{i-1} := (\mathbf{Y}_{i+1:2i-1} / \mathbf{U}_{i+1:2i-1}^\perp) (\mathbf{W}_{0:i} / \mathbf{U}_{i+1:2i-1}^\perp)^\dagger \mathbf{W}_{0:i} \quad (2.66)$$

$$= \mathcal{O}_{i-1} \mathbf{X}_{i+1} \quad (2.67)$$

y que

$$\mathcal{O}_{i-1} = \underline{\mathcal{O}}_i \quad (2.68)$$

donde $\underline{\mathcal{O}}_i$ denota a la matriz \mathcal{O}_i sin las últimas m filas. Así, \mathbf{X}_{i+1} puede ser obtenida como

$$\mathbf{X}_{i+1} = \mathcal{O}_{i-1}^\dagger \mathcal{O}_{i-1} \quad (2.69)$$

Finalmente, utilizando los datos medidos

$$\mathbf{U}_{i:i} = \begin{pmatrix} \mathbf{u}_i & \mathbf{u}_{i+1} & \cdots & \mathbf{u}_{i+j-2} & \mathbf{u}_{i+j-1} \end{pmatrix} \quad (2.70)$$

$$\mathbf{Y}_{i:i} = \begin{pmatrix} \mathbf{y}_i & \mathbf{y}_{i+1} & \cdots & \mathbf{y}_{i+j-2} & \mathbf{y}_{i+j-1} \end{pmatrix} \quad (2.71)$$

junto a las secuencias de estados

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_i & \mathbf{x}_{i+1} & \cdots & \mathbf{x}_{i+j-2} & \mathbf{x}_{i+j-1} \end{pmatrix} \quad (2.72)$$

$$\mathbf{X}_{i+1} = \begin{pmatrix} \mathbf{x}_{i+1} & \mathbf{x}_{i+2} & \cdots & \mathbf{x}_{i+j-1} & \mathbf{x}_{i+j} \end{pmatrix} \quad (2.73)$$

obtenidas a partir de \mathbf{Y}_f , \mathbf{U}_f y \mathbf{W}_p , es posible obtener \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{D} con la ec. (2.46) donde la suma va desde $t = i$ a $t = i + j - 1$.

En MATLAB, el comando utilizado para estimar un modelo en variables de estado mediante N4SID es `n4sid()`.

2.4.3. Métodos de predicción del error

En general, los métodos de subespacios son simples de implementar, pues sólo utilizan técnicas de álgebra lineal como SVD y cálculos de pseudo-inversas. Además, no necesitan de una parametrización y se realizan en un sólo paso (esto es, no son iterativos). Por lo anterior, y a pesar de que no proveen una *solución óptima*, estos métodos son muy usados en la identificación de sistemas.

Sin embargo, la técnica de identificación de sistemas más importante son los *Métodos de Predicción del Error* (o *PEM*, por sus siglas en inglés) [33, 34]. A diferencia de los métodos de subespacios, PEM es un método paramétrico que funciona sobre un marco de trabajo basado bajo supuestos de linealidad y Gaussianidad en máxima verosimilitud. En general, estos métodos PEM involucran un problema de optimización, el cual generalmente es no-lineal, por lo que se necesitan de métodos numéricos e iterativos para encontrar el parámetro a estimar.

Para comprender cómo funciona PEM, consideremos nuevamente el siguiente sistema

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{K}\mathbf{e}_t \quad (2.74)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{e}_t. \quad (2.75)$$

Ahora supondremos que las matrices dependen de un parámetro $\boldsymbol{\theta} \in \mathbb{R}^d$, donde d es el número de parámetros utilizados, y que contamos con el siguiente conjunto de datos disponible

$$Z^N = \{\mathbf{y}_t, \mathbf{u}_t | t = 1, \dots, N\}. \quad (2.76)$$

Así, la estructura del modelo queda como

$$\mathcal{M} : \begin{aligned} \hat{\mathbf{x}}_{t+1} &= \mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{x}}_t + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_t + \mathbf{K}(\boldsymbol{\theta})\mathbf{e}_t \\ \mathbf{y}_t &= \mathbf{C}(\boldsymbol{\theta})\hat{\mathbf{x}}_t + \mathbf{D}(\boldsymbol{\theta})\mathbf{u}_t + \mathbf{e}_t. \end{aligned} \quad (2.77)$$

Utilizando dicha estructura de modelo, es posible *predecir* la salida \mathbf{y}_t utilizando la información disponible sobre la entrada \mathbf{u} y salida \mathbf{y} hasta el instante $t - 1$ mediante un predictor $\hat{\mathbf{y}}_t$, el cual es la esperanza condicional de \mathbf{y}_t dada la información hasta el instante $t - 1$, esto es [33]

$$\hat{\mathbf{x}}_{t+1} = [\mathbf{A}(\boldsymbol{\theta}) - \mathbf{K}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]\hat{\mathbf{x}}_t + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_t + \mathbf{K}(\boldsymbol{\theta})\mathbf{y}_t \quad (2.78)$$

$$\hat{\mathbf{y}}_t(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})\hat{\mathbf{x}}_t + \mathbf{D}(\boldsymbol{\theta})\mathbf{u}_t \quad (2.79)$$

lo cual también puede ser escrito como

$$\hat{\mathbf{y}}_t(\boldsymbol{\theta}) = \mathbf{W}_u(q, \boldsymbol{\theta})\mathbf{u}_t + \mathbf{W}_y(q, \boldsymbol{\theta})\mathbf{y}_t \quad (2.80)$$

con

$$\mathbf{W}_u(q, \boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})[q\mathbf{I} - \mathbf{A}(\boldsymbol{\theta}) + \mathbf{K}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1}\mathbf{B}(\boldsymbol{\theta}) \quad (2.81)$$

$$\mathbf{W}_y(q, \boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})[q\mathbf{I} - \mathbf{A}(\boldsymbol{\theta}) + \mathbf{K}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})]^{-1}\mathbf{K}(\boldsymbol{\theta}). \quad (2.82)$$

Luego, si definimos el *error de predicción* $\boldsymbol{\varepsilon}_t$ como

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \hat{\mathbf{y}}_t(\boldsymbol{\theta}) \quad (2.83)$$

entonces para cuantificar qué tan bien nuestro modelo predice la salida \mathbf{y}_t se utiliza el promedio cuadrático de estos errores para cada t , esto es,

$$V_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{t=1}^N \|\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\|_2^2 \quad (2.84)$$

donde $\|\cdot\|_2$ es la norma Euclidiana. Así, el problema principal que intenta resolver PEM es el de encontrar los parámetros $\boldsymbol{\theta}$ que minimizan el criterio (2.84), esto es,

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\theta}} V_N(\boldsymbol{\theta}). \quad (2.85)$$

La tarea de minimizar la ec. (2.84) es no lineal en los parámetros $\boldsymbol{\theta}$, por lo que se hace necesario utilizar métodos numéricos e iterativos para resolver la ec. (2.85). Por ejemplo, métodos bien conocidos como el del *gradiente descendiente* o el *método de Newton* [42, 43], los cuales iteran, respectivamente, sobre

$$\hat{\boldsymbol{\theta}}^{i+1} = \hat{\boldsymbol{\theta}}^i - \boldsymbol{\mu}_i \left. \frac{\partial V_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}^i} \quad (2.86)$$

$$\hat{\boldsymbol{\theta}}^{i+1} = \hat{\boldsymbol{\theta}}^i - \boldsymbol{\mathcal{H}}_i^{-1} \left. \frac{\partial V_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}^i} \quad (2.87)$$

donde $\boldsymbol{\mu}_i$ es el *paso de optimización* y $\boldsymbol{\mathcal{H}}_i$ está dada por

$$\boldsymbol{\mathcal{H}}_i = \left. \frac{\partial^2 V_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\hat{\boldsymbol{\theta}}^i}, \quad (2.88)$$

pueden ser utilizados para resolver el problema de optimización.

Una dificultad importante para un método iterativo de optimización es la existencia de un mínimo local. Para evitarlo se requiere una buena estimación inicial $\boldsymbol{\theta}_0$ con la cual inicializar el algoritmo iterativo. Un buen valor inicial, además de permitir alcanzar el óptimo global, usualmente también mejora la velocidad de convergencia. Una herramienta que nos entrega de manera simple estimaciones de $\boldsymbol{\theta}$ son los métodos de subespacios mencionados en las secciones anteriores. Así, los métodos de subespacios usualmente son utilizados como inicialización de PEM.

Un último comentario respecto a PEM es que es posible agregar pesos o factores a la ec. (2.84) de tal manera de que los errores se minimicen específicamente en ciertas bandas de frecuencias, logrando que el modelo obtenido logre una mejor aproximación al sistema en dichas bandas. Además, es posible agregar otras restricciones al problema como, por ejemplo, minimizar la ec. (2.84) sujeto a que el sistema sea estable, esto es, que todos los valores propios de \mathbf{A} estén dentro del círculo unitario.

En MATLAB, el comando `pem()` implementa el método PEM.

2.5. Reducción de modelos

Recordemos de la sección anterior que un sistema puede describirse mediante modelos de entrada-salida, como lo son las funciones de transferencia, y mediante modelos internos, como lo son los modelos en variables de estado. Vimos también que como ambos modelos describen el mismo sistema, entonces existe una relación entre ambas representaciones. Así, si consideramos el siguiente sistema determinístico

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \quad (2.89)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \quad (2.90)$$

entonces la relación entre⁵ $\mathbf{G}(q)$ y $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ está dada por

$$\mathbf{G}(q) = \mathbf{C}(q\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \quad (2.91)$$

donde, en esta sección, se utilizará la siguiente notación compacta

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right] := \mathbf{C}(q\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}. \quad (2.92)$$

Si ahora consideramos una matriz no singular $\mathbf{T} \in \mathbb{R}^n$, donde n es el orden del sistema, y realizamos la transformación $\bar{\mathbf{x}}_t = \mathbf{T}\mathbf{x}_t$ entonces una nueva descripción en variables de estado es obtenida con $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}}) = (\mathbf{T}\mathbf{A}\mathbf{T}^{-1}, \mathbf{T}\mathbf{B}, \mathbf{C}\mathbf{T}^{-1}, \mathbf{D})$ donde, además,

$$\mathbf{G}(q) = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right] = \left[\begin{array}{c|c} \bar{\mathbf{A}} & \bar{\mathbf{B}} \\ \hline \bar{\mathbf{C}} & \bar{\mathbf{D}} \end{array} \right]. \quad (2.93)$$

Por la igualdad anterior, es claro de que existen infinitas maneras de representar un sistema en variables de estado que posean la misma característica entrada-salida. Así, a un modelo en variables de estado $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ tal que

$$\mathbf{G}(q) = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right] \quad (2.94)$$

le llamamos una *realización* de $\mathbf{G}(q)$. Por supuesto, para una función $\mathbf{G}(q)$ podrían existir realizaciones de distintos órdenes, por lo que es importante la siguiente definición.

Definición 2.3 (Realización mínima [44]). Una realización en variables de estado $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ de $\mathbf{G}(q)$ se dice que es una *realización mínima* de $\mathbf{G}(q)$ si \mathbf{A} tiene la menor dimensión posible.

Con lo anterior, el problema de *Reducción de Modelos* puede entenderse como el de encontrar la realización mínima de una función $\mathbf{G}(q)$ o, equivalentemente, encontrar la representación en variables de estado de menor orden que describa una relación de entrada-salida dada. Sin embargo, la reducción de modelos va un poco más allá, pues considera

⁵En el caso MIMO la función de transferencia en realidad es una matriz de transferencia.

el problema más general de cómo aproximar un sistema mediante una representación en variables de estado cuyo orden sea menor al de la realización mínima. Para realizar lo anterior, la idea clave será eliminar los estados de (2.89)-(2.90) que menos aporten información a la relación entrada-salida. Antes de continuar, es necesario definir algunos conceptos fundamentales de la teoría de variables de estado.

Definición 2.4 (Controlabilidad [20]). El sistema dinámico descrito por (2.89)-(2.90) o el par (\mathbf{A}, \mathbf{B}) se dice *controlable* si, para cualquier estado inicial \mathbf{x}_0 , existe una secuencia de la entrada \mathbf{u} tal que $\mathbf{x}_k = \mathbf{0}$ para algún k .

Definición 2.5 (Observabilidad [20]). El sistema dinámico descrito por (2.89)-(2.90) o el par (\mathbf{C}, \mathbf{A}) se dice *observable* si, para algún k , el estado inicial \mathbf{x}_0 , puede ser determinado a partir del conocimiento de la entrada \mathbf{u} y la salida \mathbf{y} para $t = 0, \dots, k - 1$.

En términos simples, estos conceptos relacionan la entrada \mathbf{u} (controlabilidad) y la salida \mathbf{y} (observabilidad) con el estado \mathbf{x} . En específico, el primero nos dice si la entrada del sistema puede llevar todos los estados al origen, mientras que el segundo nos dice si los estados del sistema aparecen en la salida.

Definición 2.6 (Modos controlables y observables [44]). Sea λ un valor propio de \mathbf{A} o, equivalentemente, un modo del sistema. Entonces el modo λ se dice que es controlable (observable) si $\mathbf{x}^\top \mathbf{B} \neq \mathbf{0}$ ($\mathbf{C}\mathbf{x} \neq \mathbf{0}$) para *todos* los vectores propios por la izquierda (derecha) de \mathbf{A} asociados a λ , esto es, $\mathbf{x}^\top \mathbf{A} = \lambda \mathbf{x}^\top$ ($\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$) y $\mathbf{x} \neq \mathbf{0}$. De lo contrario, el modo se dice que es no controlable (no observable).

De la definición anterior, se sigue que un sistema es controlable (observable) si, y sólo si, cada modo es controlable (observable). Un modo que no es controlable no será afectado por la entrada, mientras que un modo que no es observable no afectará a la salida. Por lo anterior, estos modos sólo contienen información interna del sistema o, equivalentemente, no entregan información sobre la relación entrada-salida. Así, si tuviéramos que *quitar estados* para disminuir el orden del modelo, lo mejor sería eliminar aquellos relacionados a estos modos no observables y/o no controlables. Es más, si pudiésemos cuantificar qué tan controlable u observable es un modo, entonces incluso podríamos eliminar aquellos que son muy poco controlables y/o observables sin perder tanta información de entrada-salida. Afortunadamente, existen dos cuantificadores de observabilidad y controlabilidad, los cuales se definen a continuación.

Definición 2.7 (Gramiano de controlabilidad y observabilidad [40]). Sea una realización dada por $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ con \mathbf{A} estable. Considere las dos ecuaciones de Lyapunov definidas por

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top \quad (2.95)$$

$$\mathbf{Q} = \mathbf{A}^\top \mathbf{Q}\mathbf{A} + \mathbf{C}\mathbf{C}^\top \quad (2.96)$$

Entonces, las soluciones \mathbf{P} y \mathbf{Q} son llamadas, respectivamente, *Gramiano de controlabilidad* y *Gramiano de observabilidad*, donde además estas matrices son semidefinidas positivas. También, las raíces cuadradas de los valores propios de $\mathbf{P}\mathbf{Q}$ son llamados los *valores singulares de Hankel* de $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

Los Gramianos permiten cuantificar la controlabilidad y observabilidad de los modos del sistema. En específico, los valores propios del Gramiano de observabilidad describen cómo

el estado inicial \mathbf{x}_0 influye en la salida \mathbf{y}_t cuando $\mathbf{u}_t = 0$, mientras que el Gramiano de controlabilidad describe cómo la entrada \mathbf{u} influye en el vector de estado \mathbf{x} [44].

Es importante mencionar que al transformar una representación de estado con $\bar{\mathbf{x}}_t = \mathbf{T}\mathbf{x}_t$, los Gramianos cambian:

$$\bar{\mathbf{P}} = \mathbf{T}^{-1}\mathbf{P}\mathbf{T}^{-\top} \quad (2.97)$$

$$\bar{\mathbf{Q}} = \mathbf{T}^{\top}\mathbf{Q}\mathbf{T}, \quad (2.98)$$

lo que es equivalente a decir que los Gramianos dependen de la realización del sistema. El problema de esto es que pueden existir realizaciones donde el Gramiano de observabilidad nos indique que un modo es *altamente observable* mientras que el Gramiano de controlabilidad nos indica que dicho modo es *altamente no controlable*, por lo que no sabríamos qué estados eliminar para reducir el orden del sistema. Para evitar este problema, se usa una realización para la cual los Gramianos de observabilidad y controlabilidad son iguales, la cual es llamada *realización balanceada*.

Definición 2.8 (Realización balanceada [40]). Sea $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ una realización mínima. Entonces es llamada *realización balanceada* si las siguientes condiciones se cumplen:

- (I) La matriz \mathbf{A} es estable, esto es⁶, $\rho(\mathbf{A}) < 1$.
- (II) Los Gramianos \mathbf{P} y \mathbf{Q} son iguales y diagonales, esto es, existe una matriz diagonal

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad (2.99)$$

con $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$ y $\mathbf{\Sigma}$ satisfaciendo

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^{\top} + \mathbf{B}\mathbf{B}^{\top} \quad (2.100)$$

$$\mathbf{\Sigma} = \mathbf{A}^{\top}\mathbf{\Sigma}\mathbf{A} + \mathbf{C}\mathbf{C}^{\top} \quad (2.101)$$

Note que $\sigma_1, \sigma_2, \dots, \sigma_n$ son los valores singulares de Hankel de $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

Así, si obtenemos la matriz de cambio de base \mathbf{T} que convierte nuestra realización en una realización balanceada, entonces esta nueva realización nos servirá para cuantificar la observabilidad y controlabilidad al mismo tiempo, en específico, mediante los valores singulares de Hankel $\sigma_1, \sigma_2, \dots, \sigma_n$. Con lo anterior, los valores singulares de menor valor nos indicarán qué modos son los menos observables y controlables y que, por tanto, podemos eliminar manteniendo una buena aproximación de la relación entrada-salida [45].

Lema 2.3. *Supongamos que $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ es una realización balanceada con \mathbf{A} estable. Luego, particionemos $\mathbf{\Sigma}$ en $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ y $\mathbf{\Sigma}_2 = \text{diag}(\sigma_{r+1}, \sigma_2, \dots, \sigma_n)$ y, consecuentemente, escribamos \mathbf{A} , \mathbf{B} y \mathbf{C} como*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{pmatrix} \quad (2.102)$$

⁶Aquí $\rho(\mathbf{A})$ es el radio espectral de la matriz $\mathbf{A} \in \mathbb{R}^{n \times n}$ el cual es igual al máximo valor singular en valor absoluto.

donde $\mathbf{A}_1 \in \mathbb{R}^{r \times r}$, $\mathbf{B}_1 \in \mathbb{R}^{r \times m}$ y $\mathbf{C}_1 \in \mathbb{R}^{p \times r}$. Definiendo un modelo de orden reducido

$$\mathbf{G}_r(q) = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{B}_1 \\ \hline \mathbf{C}_1 & \mathbf{D} \end{array} \right] \quad (2.103)$$

entonces, se cumple lo siguiente:

- (I) El modelo $\mathbf{G}_r(q)$ es estable.
- (II) Si $\sigma_r > \sigma_{r+1}$, entonces $\mathbf{G}_r(q)$ es una realización mínima.
- (III) Para cualquier $r = 1, 2, \dots, n-1$, se cumplen la siguiente cota

$$\|\mathbf{G}(e^{j\omega}) - \mathbf{G}_r(e^{j\omega})\|_\infty \leq 2(\sigma_{r+1} + \sigma_{r+2} + \dots + \sigma_n). \quad (2.104)$$

Demostración. Ver [40]. ■

Al lema anterior se le conoce como *Reducción de Modelo mediante Truncamiento Balanceado* y nos dice que es posible considerar los r valores singulares de mayor valor y, además, nos proporciona una cota acerca del error de aproximación en el dominio de la frecuencia. Es importante mencionar, que al igual que en PEM, es posible realizar una reducción de modelo utilizando pesos de tal manera que el error de aproximación sea menor en ciertas bandas de frecuencia.

Finalmente, sólo nos queda obtener la matriz de transformación \mathbf{T} para, dado un sistema descrito en variables de estado, obtener su realización balanceada. En [46], se describe un algoritmo para obtener la matriz \mathbf{T} dada una realización mínima $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ de un sistema estable. Primero, se utiliza la factorización de Cholesky para obtener

$$\mathbf{Q} = \mathbf{R}^\top \mathbf{R} \quad (2.105)$$

con $\mathbf{R} \in \mathbb{R}^{n \times n}$ y, luego, se considera la SVD

$$\mathbf{R}\mathbf{R}^\top = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^\top \quad (2.106)$$

con $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_n$ y la matriz diagonal $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$. Luego, la matriz de transformación está dada por

$$\mathbf{T} = \mathbf{R}^{-1} \mathbf{U} \mathbf{\Sigma}^{1/2}. \quad (2.107)$$

En MATLAB, el comando utilizado para obtener un modelo de orden reducido mediante truncamiento balanceado es `balred()`.

2.6. Filtraje y suavizamiento Bayesiano

El término *filtraje* se refiere a la acción de quitarle a un elemento ciertas cosas que consideramos como impuras, con el fin de obtener un elemento más puro o limpio. En un contexto de señales y sistemas, lo anterior se puede entender como *limpiar* señales mediante la eliminación del ruido que ésta contenga. Sin embargo, el concepto de filtraje es más amplio, pues se refiere a recuperar una señal \mathbf{x} mediante la medición de otra señal \mathbf{y} , la cual generalmente contiene ruido y está relacionada con \mathbf{x} .

Lo anterior se puede entender mejor si como escenario de dichas señales tenemos un sistema representado mediante el uso de espacio de estados. Así, consideremos el siguiente sistema descrito en variables de estado

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t \quad (2.108)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t. \quad (2.109)$$

El estado del sistema \mathbf{x}_t contiene la información mínima para describir las dinámicas del sistema y usualmente no es posible medirlo. Por otro lado, la salida del sistema \mathbf{y}_t sí es medible y contiene información de \mathbf{x}_t . Así, en este contexto, filtraje se refiere a la estimación de \mathbf{x}_t mediante el conocimiento de \mathbf{y}_t . Es claro que este sistema evoluciona y podremos ir coleccionando el valor de la salida en instantes consecutivos de tiempo, esto es, podemos suponer que tendremos disponible $\mathbf{y}_{1:k} := \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$. Es con la información anterior que realmente nos interesa estimar \mathbf{x}_t , donde podemos distinguir tres casos importantes de estimación [47]:

1. Si $t = k$, se denomina **filtraje** y se estima \mathbf{x}_t con información de \mathbf{y}_k anterior al instante t (inclusive t).
2. Si $t < k$, se denomina **suavizamiento** y se estima \mathbf{x}_t con información de \mathbf{y}_k anterior y posterior al instante t .
3. Si $t > k$, se denomina **predicción** y se estima o *predice* \mathbf{x}_t con información de \mathbf{y}_k anterior al instante t .

Las técnicas de suavizamiento entregan mejores estimaciones que las técnicas de filtraje, pues se hace un mayor uso de información, esto es, mientras que con filtraje se usa la información que existe hasta el instante actual, con suavizamiento se utiliza además información futura para mejorar las estimaciones [48]. Una desventaja del suavizamiento es que las estimaciones no se pueden realizar de manera *online*, lo que significa que no es posible estimar \mathbf{x}_t en t debido a que se necesita información posterior a t .

Una cuestión importante al momento de estimar una señal es tener una medida de qué tan buena es la estimación. Esto es, si suponemos que $\hat{\mathbf{x}}_t$ es una estimación de \mathbf{x}_t , se requiere una medida para cuantificar qué tanto $\hat{\mathbf{x}}_t$ se parece a \mathbf{x}_t . Con lo anterior nace el concepto de *filtraje óptimo*. Por ejemplo, el filtro de Wiener es una solución en el dominio de la frecuencia al problema de filtraje óptimo de señales estacionarias Gaussianas [49]. Por otro lado, en [50, 51], Kalman describe la solución recursiva al problema de filtraje óptimo lineal en un sistema de tiempo discreto descrito en variables de estado [52]. Por su parte, el enfoque *Bayesiano* trata el problema del filtraje óptimo desde un punto de vista probabilístico, donde se hace un extensivo uso del teorema de Bayes y los conceptos de *información a-priori* y *a-posteriori* [53]. Como veremos más adelante, usando el enfoque Bayesiano es posible llegar a la misma solución de Kalman.

Como se dijo anteriormente, para realizar filtraje óptimo necesitamos una métrica que cuantifique la calidad de la estimación. Así, una métrica utilizada en la literatura es la varianza del error de estimación:

$$\mathbb{E}\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\top | \mathbf{y}\}. \quad (2.110)$$

Teorema 2.4 (Estimación de Mínima Varianza). Sean $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^m$ dos variables aleatorias conjuntamente distribuidas. Sea $\mathbf{f}(\mathbf{y}) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ cualquier función dependiente

de \mathbf{y} y sea $\mathbf{J}(\mathbf{f}(\mathbf{y}))$ el funcional de costo definido como

$$\mathbf{J}(\mathbf{f}(\mathbf{y})) = \mathbb{E}\{(\mathbf{x} - \mathbf{f}(\mathbf{y}))(\mathbf{x} - \mathbf{f}(\mathbf{y}))^\top | \mathbf{y}\} \quad (2.111)$$

Entonces, la estimación de mínima varianza denotada por $\hat{\mathbf{x}}$, esto es,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{f}(\mathbf{y})} \mathbf{J}(\mathbf{f}(\mathbf{y})) \quad (2.112)$$

está dada por $\hat{\mathbf{x}} = \mathbb{E}\{\mathbf{x} | \mathbf{y}\}$.

Demostración. Ver [54]. ■

En el contexto de variables de estado que hemos planteado, el teorema anterior nos dice que si queremos obtener la estimación del estado $\hat{\mathbf{x}}_t$ que minimiza la varianza del error de estimación dadas las mediciones $\mathbf{y}_{1:k}$, entonces debemos calcular $\mathbb{E}\{\mathbf{x}_t | \mathbf{y}_{1:k}\}$. El cálculo anterior implica conocer la función de probabilidad $p(\mathbf{x}_t | \mathbf{y}_{1:k})$, la cual generalmente es difícil de encontrar. Sin embargo, mediante el teorema de Bayes, dicha función de probabilidad puede escribirse como:

$$p(\mathbf{x}_t | \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_{1:k} | \mathbf{x}_t)p(\mathbf{x}_t)}{p(\mathbf{y}_{1:k})} \quad (2.113)$$

donde los términos del lado derecho, en general, son más fáciles de encontrar.

La resolución del problema anterior, mediante la perspectiva Bayesiana, se realiza sobre modelos en espacio de estados probabilísticos generales, los cuales tienen la siguiente forma.

Definición 2.9 (Modelo en espacio de estados probabilístico [47]). Un modelo en espacio de estados probabilístico o modelo de filtraje consiste en una secuencia de distribuciones de probabilidad condicional:

$$\mathbf{x}_{t+1} | \mathbf{x}_t \sim p(\mathbf{x}_{t+1} | \mathbf{x}_t) \quad (2.114)$$

$$\mathbf{y}_t | \mathbf{x}_t \sim p(\mathbf{y}_t | \mathbf{x}_t) \quad (2.115)$$

para $t = 1, 2, \dots$, donde

- $\mathbf{x}_t \in \mathbb{R}^n$ es el estado del sistema en el tiempo t .
- $\mathbf{y}_t \in \mathbb{R}^m$ es la medición en el tiempo t .
- $p(\mathbf{x}_{t+1} | \mathbf{x}_t)$ es el modelo que describe la dinámica del sistema.
- $p(\mathbf{y}_t | \mathbf{x}_t)$ es el modelo de medición.

El modelo se asume que es de Markov, esto quiere decir que cumple las siguientes propiedades:

1. Propiedad de estados de Markov

- $p(\mathbf{x}_t | \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$
- $p(\mathbf{x}_{t-1} | \mathbf{x}_{t:T}, \mathbf{y}_{t:T}) = p(\mathbf{x}_{t-1} | \mathbf{x}_t)$

2. Independencia condicional de las mediciones

- $p(\mathbf{y}_t | \mathbf{x}_{1:t}, \mathbf{y}_{1:t-1}) = p(\mathbf{y}_t | \mathbf{x}_t)$

En las próximas sub-secciones se resuelven los problemas de filtraje y suavizamiento Bayesiano, además de obtener la solución cerrada en el caso donde los modelos dinámico y de medición son lineales y Gaussianos. Dichas soluciones cerradas son conocidas como *filtro de Kalman* y *suavizador de Kalman* (o *RTS*), respectivamente.

2.6.1. Filtraje Bayesiano

El propósito del filtraje Bayesiano es calcular la distribución del estado \mathbf{x}_t en cada instante de tiempo t dado el historial de mediciones hasta el instante t , esto es:

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) \quad (2.116)$$

Las ecuaciones fundamentales de la teoría de filtraje Bayesiano están dadas por el siguiente teorema.

Teorema 2.5 (Ecuaciones de filtraje Bayesiano). *Las ecuaciones recursivas para calcular la distribución de Filtrado $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ y la distribución de Predicción $p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})$ en cada instante de tiempo t están dadas por:*

- **Filtrado:** Dada la medición \mathbf{y}_t en el instante t , la distribución del estado \mathbf{x}_t puede ser calculado mediante la regla de Bayes

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})} \quad (2.117)$$

donde $p(\mathbf{y}_t | \mathbf{y}_{1:t-1})$ es una constante de normalización que puede ser calculada como

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) d\mathbf{x}_t \quad (2.118)$$

- **Predicción:** La distribución de Predicción del estado \mathbf{x}_{t+1} en el instante t puede ser calculada mediante la ecuación de Chapman-Kolmogorov

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t}) d\mathbf{x}_t \quad (2.119)$$

La recursión comienza a partir de la distribución a priori $p(\mathbf{x}_0 | \mathbf{y}_{-1}) = p(\mathbf{x}_0)$.

Demostración. Ver [47]. ■

Como se ha dicho anteriormente, el filtro de Kalman [50] es la solución cerrada a las ecuaciones de filtraje Bayesiano cuando el sistema es lineal y los ruidos de proceso y de salida son Gaussianos. Así, consideremos nuevamente el sistema (2.108)-(2.109):

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t \quad (2.120)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \quad (2.121)$$

donde $\mathbf{w}_t \sim \mathcal{N}_{\mathbf{w}_t}(\mathbf{0}, \mathbf{Q})$, $\mathbf{v}_t \sim \mathcal{N}_{\mathbf{v}_t}(\mathbf{0}, \mathbf{R})$ y $\mathbf{x}_0 \sim \mathcal{N}_{\mathbf{x}_0}(\boldsymbol{\mu}_0, \mathbf{P}_0)$. Con lo anterior, tenemos el siguiente teorema.

Teorema 2.6 (Filtro de Kalman). *Las ecuaciones de filtraje Bayesiano para el modelo lineal (2.120)-(2.121) pueden ser evaluadas en una forma cerrada y las distribuciones resultantes son Gaussianas:*

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \mathcal{N}_{\mathbf{x}_{t+1}}(\hat{\mathbf{x}}_{t+1|t}, \mathbf{P}_{t+1|t}) \quad (2.122)$$

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) = \mathcal{N}_{\mathbf{x}_t}(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \quad (2.123)$$

Los parámetros de las distribuciones anteriores pueden ser calculadas con los siguientes pasos de Filtrado y Predicción:

- *Filtrado:*

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{C}^\top (\mathbf{C} \mathbf{P}_{t|t-1} \mathbf{C}^\top + \mathbf{R})^{-1} \quad (2.124)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{C} \hat{\mathbf{x}}_{t|t-1}) \quad (2.125)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{C}) \mathbf{P}_{t|t-1} \quad (2.126)$$

- *Predicción:*

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{A} \hat{\mathbf{x}}_{t|t} \quad (2.127)$$

$$\mathbf{P}_{t+1|t} = \mathbf{A} \mathbf{P}_{t|t} \mathbf{A}^\top + \mathbf{Q} \quad (2.128)$$

donde la recursión comienza a partir de $\hat{\mathbf{x}}_{0|-} = \boldsymbol{\mu}_0$ y $\mathbf{P}_{0|-} = \mathbf{P}_0$.

Demostración. Ver [47]. ■

El sistema (2.120)-(2.121) junto a las ecuaciones del filtro de Kalman aquí mostradas consideran el caso más simple, esto es, no consideran ni entrada \mathbf{u}_t al sistema ni matrices variantes en el tiempo ni correlación cruzada \mathbf{S} entre los ruidos \mathbf{w}_t y \mathbf{v}_t . Lo anterior es relativamente fácil de introducir. Por ejemplo, en el caso de considerar matrices variantes en el tiempo (por ejemplo, considerar \mathbf{A}_t en vez de \mathbf{A}), las ecuaciones siguen siendo exactamente iguales (con la diferencia de agregar la dependencia del tiempo en las matrices). Por otro lado, para introducir la correlación cruzada \mathbf{S} basta con realizar los siguientes cambios a las ecuaciones anteriores [53]:

$$\mathbf{A} \rightarrow \mathbf{A} - \mathbf{S} \mathbf{R}^{-1} \mathbf{C} \quad (2.129)$$

$$\mathbf{Q} \rightarrow \mathbf{Q} - \mathbf{S} \mathbf{R}^{-1} \mathbf{S}^\top \quad (2.130)$$

$$\hat{\mathbf{x}}_{t+1|t} \rightarrow \mathbf{A} \hat{\mathbf{x}}_{t|t} + \mathbf{S} \mathbf{R}^{-1} \mathbf{y}_t \quad (2.131)$$

2.6.2. Suavizamiento Bayesiano

El propósito del suavizamiento Bayesiano es calcular la distribución del estado \mathbf{x}_t luego de haber recibido las mediciones hasta el instante T , con $T > t$, esto es:

$$p(\mathbf{x}_t | \mathbf{y}_{1:T}). \quad (2.132)$$

Así, el siguiente teorema nos entrega las ecuaciones para calcular dicha distribución.

Teorema 2.7 (Ecuaciones óptimas de suavizamiento Bayesiano). *Las ecuaciones recursivas hacia atrás (suavizador Bayesiano) para calcular las distribuciones suavizadas $p(\mathbf{x}_t | \mathbf{y}_{1:T})$ para cualquier $t < T$ están dadas por las siguientes ecuaciones de suavizamiento Bayesiano de intervalo fijo:*

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t}) d\mathbf{x}_t \quad (2.133)$$

$$p(\mathbf{x}_t | \mathbf{y}_{1:T}) = p(\mathbf{x}_t | \mathbf{y}_{1:t}) \int \left[\frac{p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T})}{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})} \right] d\mathbf{x}_{t+1} \quad (2.134)$$

donde $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ es la distribución de Filtrado y $p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})$ es la distribución de Predicción.

Demostración. Ver [47]. ■

El *suavizador RTS* [55] (nombrado así por sus autores Rauch, Tung y Striebel), que también es llamado *suavizador de Kalman*, es la solución cerrada a las ecuaciones de suavizamiento Bayesiano para el caso donde el sistema es lineal y los ruidos de proceso y de salida son Gaussianos. Así, el siguiente teorema nos entrega las ecuaciones que describen la manera en que se obtienen las estadísticas de las distribuciones.

Teorema 2.8 (Suavizador RTS o de Kalman). *La distribución $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ es Gaussiana:*

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = \mathcal{N}_{\mathbf{x}_t}(\hat{\mathbf{x}}_{t|T}, \mathbf{P}_{t|T}) \quad (2.135)$$

Luego, las ecuaciones recursivas hacia atrás para el suavizador RTS (el cual es de intervalo fijo) están dadas por

$$\mathbf{G}_t = \mathbf{P}_{t|t} \mathbf{A}^\top \mathbf{P}_{t+1|t}^{-1} \quad (2.136)$$

$$\hat{\mathbf{x}}_{t|T} = \hat{\mathbf{x}}_{t|t} + \mathbf{G}_t (\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}) \quad (2.137)$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{G}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{G}_t^\top \quad (2.138)$$

donde $\hat{\mathbf{x}}_{t|t}$ y $\mathbf{P}_{t|t}$ corresponden a los pasos de *Measurement Update* del filtro de Kalman, mientras que $\hat{\mathbf{x}}_{t+1|t}$ y $\mathbf{P}_{t+1|t}$ corresponden a los pasos de *Time Update*. La recursión comienza hacia adelante a partir de $\hat{\mathbf{x}}_{0|-} = \boldsymbol{\mu}_0$ y $\mathbf{P}_{0|-} = \mathbf{P}_0$ para luego devolverse hacia atrás en el instante $t = T$, con $\hat{\mathbf{x}}_{T|T}$ y $\mathbf{P}_{T|T}$, luego de haber obtenido $\hat{\mathbf{x}}_{t|t}$, $\mathbf{P}_{t|t}$, $\hat{\mathbf{x}}_{t+1|t}$ y $\mathbf{P}_{t+1|t}$ con el filtro de Kalman para $1 \leq t \leq T$.

Demostración. Ver [47]. ■

Las ecuaciones de filtraje y suavizamiento de Kalman han sido implementadas fácilmente en el software MATLAB.

MODELO BASADO EN IMPEDANCIA SUBGLOTAL DE ORDEN REDUCIDO

En este capítulo se presenta el modelo basado en impedancia subglotal, el cual es un modelo matemático que describe la relación entre la señal de un acelerómetro puesto sobre la piel del cuello y el flujo de aire que atraviesa la glotis. Este modelo puede ser descrito mediante su respuesta en frecuencia o mediante su respuesta a impulso. Sin embargo, dado que queremos implementar un suavizador de Kalman es necesario obtener su representación en variables de estado. Por lo anterior, en las secciones siguientes se utilizan distintas técnicas para obtener una representación en variables de estado de bajo orden que represente de la mejor manera posible el modelo basado en impedancia subglotal. En específico, tres técnicas diferentes son utilizadas: reducción de orden, identificación en el dominio del tiempo e identificación en el dominio de la frecuencia.

3.1. Modelo basado en impedancia subglotal

Los circuitos eléctricos lineales son sistemas que han sido estudiados por largos años. El modelamiento de éstos se realiza mediante el uso de leyes elementales que los describen, como lo son las leyes de Kirchhoff, y mediante leyes específicas que satisfacen los diferentes componentes electrónicos, como por ejemplo la ley de Ohm. Utilizando estas leyes, es posible relacionar las dos señales más importantes de un circuito eléctrico: el voltaje y la corriente. Así, un circuito eléctrico lineal queda descrito matemáticamente mediante ecuaciones diferenciales ordinarias, con lo que es posible utilizar las transformadas de Laplace y Fourier para analizar el circuito y, en específico, sus señales internas, desde una perspectiva frecuencial.

Análogamente, en un sistema donde existen flujos que atraviesan tuberías o tanques, las señales de presión y flujo satisfacen leyes matemáticamente equivalentes a las del voltaje y la corriente en un circuito eléctrico. Por lo anterior, es posible modelar un sistema de flujo mediante analogías de circuitos eléctricos, donde la presión, el flujo, la resistencia de flujo, la

compliance y la inercia equivalen, respectivamente, al voltaje, la corriente, la resistencia eléctrica, la capacitancia y la inductancia [56]. Así, si tomamos en cuenta que el sistema respiratorio del ser humano, el cual está encargado de la producción de la voz, es uno donde existen flujos de aire comandados por presiones generadas por los pulmones, entonces es posible analizarlo tal como lo haríamos con un circuito eléctrico, utilizando la transformada de Fourier junto al concepto de impedancia para analizarlo en el dominio de la frecuencia.

En efecto, utilizando analogías mecano-acústicas, principios de líneas de transmisión y descripciones fisiológicas, es posible construir un modelo acústico del sistema subglotal (parte del tracto respiratorio bajo la glotis) [22, 57–59], el cual se obtiene mediante la interconexión de pequeñas secciones tubulares de paredes rígidas que provienen desde la tráquea y terminan en los alvéolos, tal como lo muestra la Fig. 3.1.

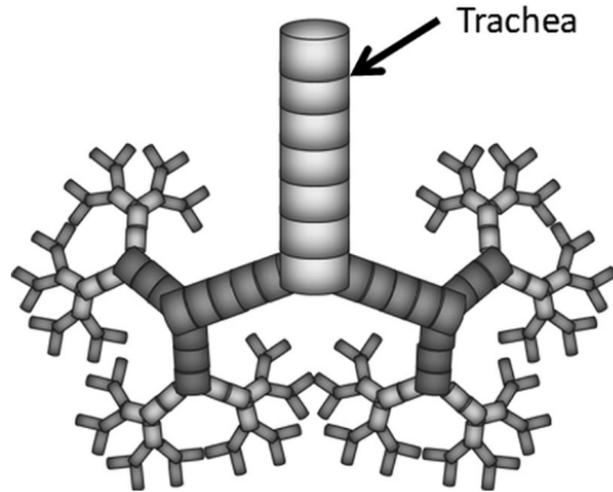


Figura 3.1: Sistema subglotal modelado como una red de tubos cilíndricos en forma de árbol invertido [59].

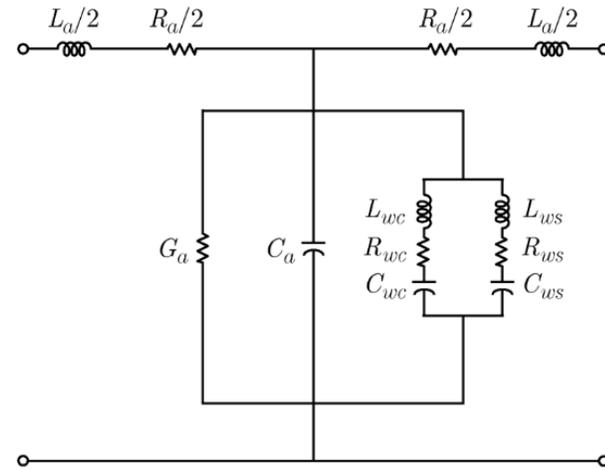
Bajo ciertas suposiciones relacionadas al largo, estos tubos quedan representados por circuitos de parámetros concentrados configurados en T, los cuales relacionan la presión acústica $P(\omega)$ con la velocidad volumétrica del flujo de aire $U(\omega)$ (donde ω es la frecuencia) [22]. En la Fig. 3.2-a se muestra una representación del circuito configurado en T, donde se incluyen parámetros relacionados al tejido cartilaginoso (L_{wc} , R_{wc} , C_{wc}), al tejido blando (L_{ws} , R_{ws} , C_{ws}), a la viscosidad del aire (R_a), a las pérdidas por conducción de calor (G_a), a la elasticidad (C_a) y a la inercia (L_a). Posteriormente, mediante el uso de impedancias, es posible representar dicho segmento como una red de dos puertos configurada en T, tal como se muestra en la Fig. 3.2-b.

Luego, la matriz de transmisión que relaciona dos secciones T consecutivas tiene la siguiente estructura:

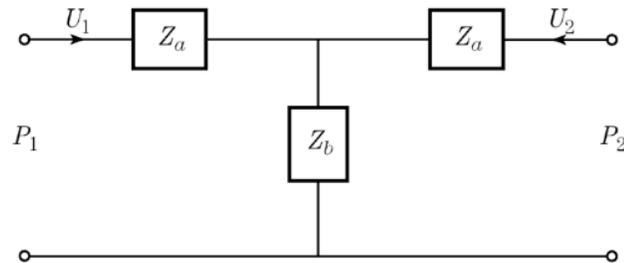
$$\begin{bmatrix} P_1(\omega) \\ U_2(\omega) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P_2(\omega) \\ -U_2(\omega) \end{bmatrix} \quad (3.1)$$

donde

$$A = (Z_a + Z_b)Z_b^{-1} \quad (3.2)$$



(a) Modelo acústico del segmento básico usado en el tracto subglotal, el cual incluye representaciones acústicas de las pérdidas, la elasticidad, la inercia y las paredes no rígidas.



(b) Representación como red de dos puertos configurada en T del segmento básico del tracto subglotal mediante el uso de impedancias equivalentes.

Figura 3.2: Representación del elemento básico usado en el modelo acústico del sistema subglotal [22].

$$B = (Z_a + 2Z_b)Z_a Z_b^{-1} \quad (3.3)$$

$$C = Z_b^{-1} \quad (3.4)$$

$$D = A. \quad (3.5)$$

En esta configuración tenemos que la respuesta en frecuencia de la función de transferencia de flujo $H(\omega) = -U_2(\omega)/U_1(\omega)$ está dada por

$$H(\omega) = \frac{1}{CZ_2(\omega) + D} \quad (3.6)$$

mientras que la impedancia de entrada $Z_1(\omega) = P_1(\omega)/U_1(\omega)$ (esto es, la impedancia vista desde la izquierda de $P_1(\omega)$) está dada por

$$Z_1(\omega) = \frac{AZ_2(\omega) + B}{CZ_2(\omega) + D} \quad (3.7)$$

donde $Z_2(\omega) = P_2(\omega)/U_2(\omega)$ denota la impedancia de carga efectiva (esto es, la impedancia que se ve hacia la derecha de las terminales de $P_2(\omega)$).

Para obtener un modelo del tracto subglotal, dichos segmentos básicos son unidos en cascada, trayendo hacia atrás la impedancia de entrada de los subtractos previos, partiendo desde la impedancia terminal del pulmón hasta terminar en la glotis. Lo anterior, da como resultado una impedancia equivalente, conocida como impedancia subglotal [57], que relaciona el flujo de aire que pasa por la glotis con el flujo de aire que sale de los pulmones.

Hasta este punto, hemos relacionado flujos de aire *internos*, pero aún no hemos incluido la relación con la aceleración sobre la piel. Para ello, es necesario conocer la relación entre la presión que existe sobre las paredes internas del tracto subglotal (a la altura del acelerómetro) y el flujo de aire que pasa por dichas paredes. En efecto, esta relación queda descrita por la impedancia de la piel del cuello $Z_{skin}(\omega)$ [60]. En este trabajo (y en [22]), dicha impedancia tiene como objetivo relacionar la señal de aceleración sobre la piel del cuello $\dot{U}_{skin}(\omega)$ con el flujo de aire justo debajo de la glotis $U_{sub}(\omega)$, además de considerar las propiedades mecánicas de la piel junto al efecto de la carga del acelerómetro [22]. En la Fig. 3.3 se muestra un esquema del tracto subglotal donde se ha incluido la impedancia $Z_{skin}(\omega)$, la cual es conectada en paralelo entre la sección *sub1*, representada por su impedancia de entrada $Z_{sub1}(\omega)$ y la sección *sub2*, representada por su impedancia de entrada $Z_{sub2}(\omega)$. Luego, en la Fig. 3.4 se muestra una representación fisiológica de dicho esquema, donde se visualiza la posición del acelerómetro junto a dos secciones del tracto subglotal, *sub1* y *sub2*, que representan las porciones de la tráquea extratorácica sobre y bajo la posición del acelerómetro, respectivamente.

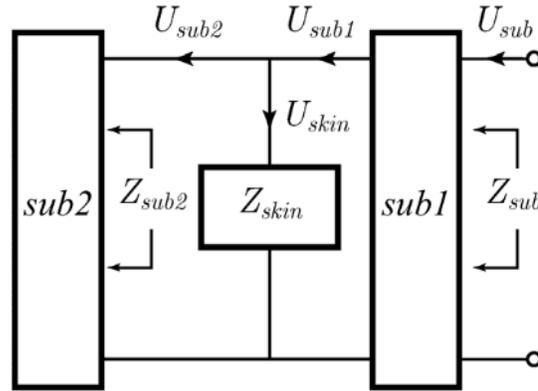


Figura 3.3: Representación del modelo del tracto subglotal donde se ha incluido la impedancia $Z_{skin}(\omega)$ para considerar los efectos de la piel y de la carga del acelerómetro [22].

Específicamente, la expresión utilizada para $Z_{skin}(\omega)$ está dada por [22]

$$Z_{skin}(\omega) = Z_m(\omega) + Z_{rad}(\omega) \quad (3.8)$$

$$Z_m(\omega) = R_m + j \left(\omega M_m - \frac{K_m}{\omega} \right) \quad (3.9)$$

$$Z_{rad}(\omega) = \frac{j\omega M_{acc}}{A_{acc}} \quad (3.10)$$

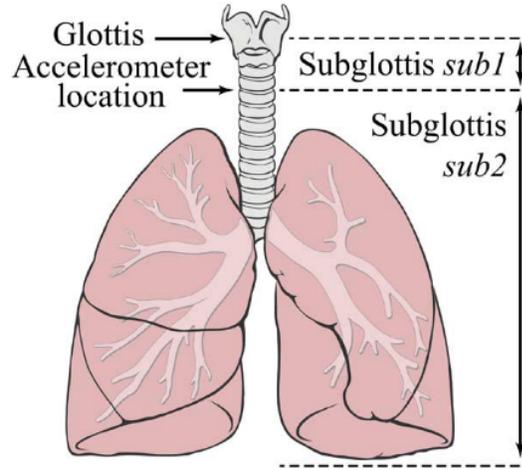


Figura 3.4: Representación fisiológica del sistema subglotal. El acelerómetro es puesto sobre la superficie de la piel que cubre la muesca supraesternal aproximadamente 5 cm bajo la glottis. Los trectos sobre y bajo esta posición son denotados como *sub1* y *sub2*, respectivamente [22].

donde $Z_m(\omega)$ representa la impedancia mecánica de la piel y cuyos parámetros R_m , M_m y K_m son, respectivamente, la resistencia, inercia y rigidez por unidad de área de la piel, mientras que $Z_{rad}(\omega)$ representa la impedancia de radiación debido a la carga del acelerómetro y cuyos parámetros M_{acc} y A_{acc} son, respectivamente, la masa y superficie del acelerómetro.

A continuación, mediante el uso de un *divisor de corrientes* sobre la Fig. 3.3, es posible obtener una expresión para la velocidad volumétrica $U_{skin}(\omega)$ que existe sobre $Z_{skin}(\omega)$, la cual está dada por

$$U_{skin}(\omega) = U_{sub1}(\omega) \frac{Z_{sub2}(\omega)}{Z_{sub2}(\omega) + Z_{skin}(\omega)} \quad (3.11)$$

donde $U_{sub1}(\omega)$ es el flujo de aire que sale de la sección *sub1*. Luego, como realmente se tienen mediciones de la aceleración, es necesario derivar $U_{skin}(\omega)$ en el dominio del tiempo para obtener dicha señal. Así, como la derivada en el dominio del tiempo se representa por una multiplicación por $j\omega$ en el dominio de la frecuencia, basta con multiplicar la expresión anterior por $j\omega$ para obtener lo que deseamos, esto es,

$$\dot{U}_{skin}(\omega) = U_{sub1}(\omega) \frac{Z_{sub2}(\omega)}{Z_{sub2}(\omega) + Z_{skin}(\omega)} j\omega. \quad (3.12)$$

Finalmente, la relación entre la señal de aceleración $\dot{U}_{skin}(\omega)$ y el flujo de aire justo bajo la glottis $U_{sub}(\omega)$, denotada como $T_{skin}(\omega)$, está dada por

$$T_{skin}(\omega) = \frac{\dot{U}_{skin}(\omega)}{U_{sub}(\omega)} = \frac{1}{U_{sub}(\omega)} \cdot \left(U_{sub1}(\omega) \frac{Z_{sub2}(\omega)}{Z_{sub2}(\omega) + Z_{skin}(\omega)} j\omega \right) \quad (3.13)$$

o, equivalentemente, por

$$T_{skin}(\omega) = \frac{H_{sub1}(\omega) \cdot Z_{sub2}(\omega) \cdot j\omega}{Z_{sub2}(\omega) + Z_{skin}(\omega)} \quad (3.14)$$

donde $H_{sub1}(\omega) = U_{sub1}(\omega)/U_{sub}(\omega)$ es la respuesta en frecuencia de la función de transferencia de la sección subglotal *sub1*, la cual va desde la glotis hasta la ubicación del acelerómetro y donde $Z_{sub2}(\omega)$, como ya se ha dicho, es la impedancia de entrada vista hacia la sección subglotal *sub2*, la cual va desde la ubicación del acelerómetro hasta los pulmones (ver Fig. 3.3 y 3.4). Así, al modelo que representa $T_{skin}(\omega)$ se le conoce como *modelo basado en impedancia subglotal (IB subglottal model, en inglés)* [22].

De la expresión (3.14), se debe decir que $H_{sub1}(\omega)$ y $Z_{sub2}(\omega)$ son obtenidos tal como se explicó al principio de este capítulo. En específico, los segmentos básicos que componen las secciones correspondientes (*sub1* y *sub2*) son concatenados en serie. Así, $Z_{sub2}(\omega)$ se obtiene al ir trayendo *hacia atrás* la impedancia de entrada de los segmentos previos, partiendo de la impedancia terminal de los pulmones y aplicando sucesivamente (3.7), mientras que $H_{sub1}(\omega)$ se obtiene de manera similar mediante el uso de (3.6).

Otro punto importante que se debe mencionar, es que en (3.14) existen parámetros que definen el modelo basado en impedancia subglotal y que dependen de cada sujeto en estudio. Estos parámetros son 5: la resistencia (R_m), la inercia (M_m) y la elasticidad (K_m) de la piel, incluidos en $Z_{skin}(\omega)$ y que están relacionados a las propiedades mecánicas de la piel, y el largo de la tráquea ($L_{trachea}$) junto a la posición del acelerómetro respecto a la glotis (L_{sub1}) incluidos implícitamente en $H_{sub1}(\omega)$ y $Z_{sub2}(\omega)$. Para obtener estos parámetros se utiliza un esquema de calibración, donde se utiliza un vector de escalamiento $\mathbf{Q} = \{Q_i\}_{i=1,\dots,5}$ que permite regular los parámetros a partir de sus valores por defecto de la siguiente manera:

$$R_m = 2320 \cdot Q_1 \quad [\text{g} \cdot \text{s}^{-1} \cdot \text{cm}^{-2}] \quad (3.15)$$

$$M_m = 2.4 \cdot Q_2 \quad [\text{g} \cdot \text{cm}^{-2}] \quad (3.16)$$

$$K_m = 491000 \cdot Q_3 \quad [\text{dyn} \cdot \text{cm}^{-3}] \quad (3.17)$$

$$L_{trachea} = 10 \cdot Q_4 \quad [\text{cm}] \quad (3.18)$$

$$L_{sub1} = 5 \cdot Q_5 \quad [\text{cm}] \quad (3.19)$$

donde los parámetros por defecto del modelo son obtenidos con $\mathbf{Q} = [1, 1, 1, 1, 1]$. Cuando se desea estimar estos parámetros, para que exista significado fisiológico, se restringe a que $Q_i \in D_i$ para $i = 1, \dots, 5$ con $\mathbf{D} = \{D_i\}_{i=1,\dots,5}$ definido como

$$D_i = \begin{cases} [0.1, 20] & \text{si } i = 1, 2, 3 \\ [0.6, 1.2] & \text{si } i = 4, 5. \end{cases} \quad (3.20)$$

Así, para obtener los parámetros \mathbf{Q} que definen al modelo IB subglotal específico para un sujeto, se compara la estimación del flujo de aire glotal obtenido mediante este modelo con una señal de referencia, la cual es equivalente al flujo de aire glotal obtenido mediante filtraje inverso del flujo de aire oral, el cual es capturado mediante un neumotacógrafo o máscara de Rothenberg [14]. En efecto, una vez obtenida la estimación del flujo de aire glotal obtenida a partir de la máscara de Rothenberg, es posible utilizar un esquema de *Optimización por Enjambre de Partículas* (PSO, por sus siglas en inglés), el cual busca obtener, mediante un proceso iterativo, los parámetros \mathbf{Q}^* que optimizan una función de costo. En específico, este

esquema define cada conjunto de parámetros \mathbf{Q} como una partícula, para luego definir p partículas $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_p$ las cuales se van moviendo en el espacio de parámetros según su óptimo particular y según el óptimo grupal, hasta que todas convergen en el óptimo de la función de costo.

Con lo anterior, los parámetros \mathbf{Q}^* que definen al modelo IB subglotal de un sujeto, están dados por

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q}} f(\mathbf{Q}), \quad \text{sujeto a } \mathbf{Q} \in D \quad (3.21)$$

donde $f(\cdot) : \mathbb{R}^5 \rightarrow \mathbb{R}$ está dada por [17]

$$f(\mathbf{Q}) = \sum_{i=1}^3 w_i \cdot e_i(\mathbf{Q}) \quad (3.22)$$

con

$$\sum_{i=1}^3 w_i = 1, \quad 0 \leq w_i \leq 1 \quad (3.23)$$

y

$$e_i(\mathbf{Q}) = \frac{\sum_{t=0}^{n-1} |\Delta^{(i-1)} \tilde{u}_t^g - \Delta^{(i-1)} \hat{u}_t^g|}{\sum_{t=0}^{n-1} |\Delta^{(i-1)} \tilde{u}_t^g|} \quad (3.24)$$

donde \tilde{u}_t^g es la señal de flujo de aire glotal obtenida a partir de filtraje inverso utilizando la máscara de Rothenberg, \hat{u}_t^g es la señal de flujo de aire glotal alineada temporalmente y obtenida mediante el modelo IB subglotal, $\Delta^{(i-1)}$ es el operador derivada temporal de orden $(i-1)$, i representa el índice de la función de error correspondiente e_i y su peso w_i , y n es el número de puntos muestreados de cada señal. Adicionalmente, en este trabajo cada peso w_i fue escogido como $w_i = 1/3$.

Una última observación es el hecho de que $T_{skin}(\omega)$ presenta un cero en $\omega = 0$, por lo que su inversión (cuestión que será explicada en el próximo capítulo) genera una gran magnitud a bajas frecuencias. Por lo anterior, en [22] se corrige el valor de $|T_{skin}(\omega)|$ para bajas frecuencias al forzar $|T_{skin}(0)| = 1$.

En la Fig. 3.5 se muestra un ejemplo de la magnitud y fase de la respuesta en frecuencia $T_{skin}(\omega)$ del modelo IB subglotal para el sujeto FN001 (ver Tabla 3.1) al utilizar la vocal /a/ en el esquema de calibración de los parámetros \mathbf{Q} . Debido a que no es posible obtener una función racional de la forma (2.23) para (3.14), esta respuesta en frecuencia se ha obtenido de manera discreta, esto es, $T_{skin}(\omega)$ se ha muestreado, obteniendo $T_{skin}[k] = T_{skin}(\omega_k)$, que equivale a la transformada discreta de Fourier (DFT, por sus siglas en inglés, o FFT, algoritmo estándar para obtenerla), con $\omega_k = 2\pi k f_s / N$ para $k = 0, 1, \dots, N-1$ y donde f_s es la frecuencia de muestreo y N es el número de puntos utilizados para obtener la FFT. Aquí es importante mencionar que $T_{skin}[k]$ es forzada a ser conjugada simétrica, esto es, que cumpla $T_{skin}[k] = T_{skin}^*[N-k]$, para que al calcular la FFT inversa resulte en una secuencia real.

Por otro lado, en la Fig. 3.6 se muestra la FFT inversa (IFFT, por sus siglas en inglés) de $T_{skin}[k]$, la cual equivale a la respuesta a impulso del modelo IB subglotal. Es importante notar que la respuesta a impulso h_t es distinta de 0 para $t < 0$, por lo que podemos decir que dicha respuesta a impulso es no causal.

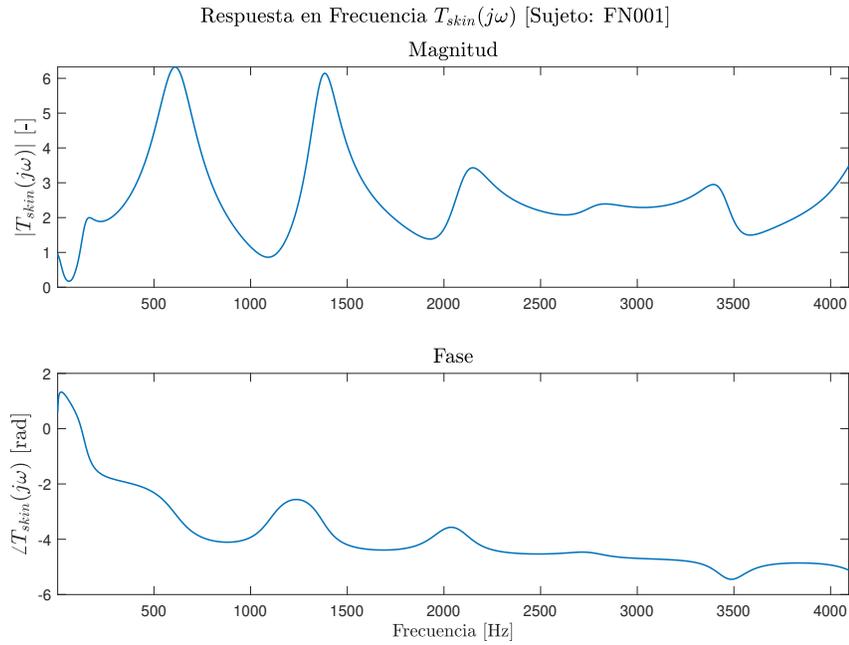


Figura 3.5: Magnitud y fase de la respuesta en frecuencia $T_{skin}(j\omega)$ del modelo IB subglotal. Los parámetros utilizados fueron los del sujeto FN001, con $f_s = 8192$ Hz y $N = 4096$, para la vocal /a/.

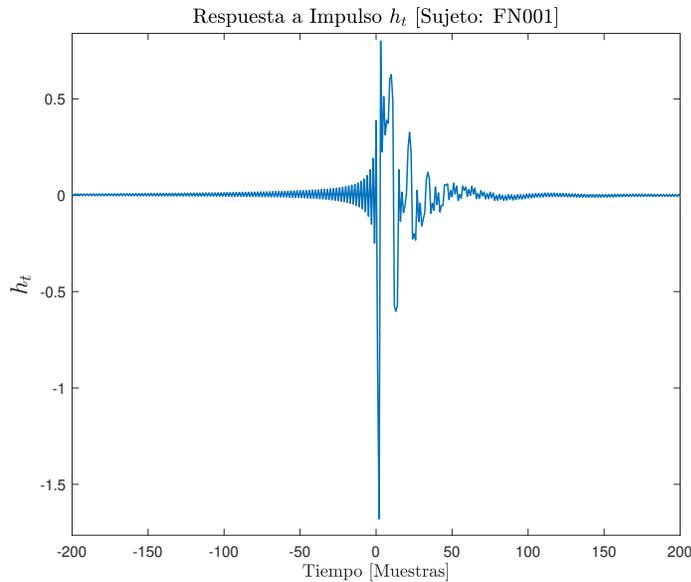


Figura 3.6: Respuesta a impulso h_t del modelo basado en impedancia subglotal. Los parámetros utilizados fueron los del sujeto FN001, con $f_s = 8192$ Hz y $N = 4096$, para la vocal /a/.

A lo largo del presente trabajo se utilizarán los datos de 6 sujetos, cuyos parámetros \mathbf{Q} , considerando la vocal /a/ en el esquema de calibración, se muestran en la Tabla 3.1. En ella FP y FN hacen referencia a pacientes con hiperfunción fonotraumática y no fonotraumática, respectivamente. Además, se utilizará una frecuencia de muestreo igual a $f_s = 8192$ Hz y la cantidad de números para obtener la FFT será de $N = 4096$.

Parámetros \mathbf{Q}					
Sujeto	$Q_{1,2,3} \in [0.1, 20]$			$Q_{4,5} \in [0.6, 1.2]$	
	Q_1	Q_2	Q_3	Q_4	Q_5
FN001	1.3993	2.7689	11.5881	1.1	0.6
FN003	0.6659	4.9677	19.7249	1.1	0.6
FN006	0.5282	2.9953	16.6386	1.2	0.8
FP003	0.5918	3.0247	16.4503	1.1	0.6
FP004	0.5626	2.3906	6.2257	1.1	0.6
FP008	0.2395	1.8631	8.4506	1.1	0.6

Tabla 3.1: Parámetros \mathbf{Q} de los sujetos analizados en este trabajo.

3.2. Representación en variables de estado del modelo IB subglotal

Como se ha dicho en la introducción de este capítulo, para ser capaces de aplicar un enfoque Bayesiano al problema actual, es necesario contar con una representación en variables de estado del modelo IB subglotal. Lo anterior es equivalente a encontrar matrices $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s, \mathbf{D}_s)$ tal que el sistema descrito por

$$\mathbf{x}_{t+1}^s = \mathbf{A}_s \mathbf{x}_t^s + \mathbf{B}_s u_t^g \quad (3.25)$$

$$y_t^a = \mathbf{C}_s \mathbf{x}_t^s + \mathbf{D}_s u_t^g \quad (3.26)$$

donde u_t^g es el flujo de aire glotal y y_t^a es la señal del acelerómetro, tenga la misma respuesta en frecuencia $T_{skin}(\omega)$ y/o la misma respuesta a impulso h_t que el modelo IB subglotal original.

Para realizar lo anterior, debemos considerar que tenemos como información tanto la respuesta en frecuencia $T_{skin}(\omega)$ como la respuesta a impulso h_t de dicho sistema (la cual equivale a la IFFT de $T_{skin}(\omega)$). Así, en las siguientes subsecciones, nos concentraremos en encontrar las matrices $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s, \mathbf{D}_s)$ a partir de información en el dominio del tiempo, esto es, utilizando h_t y a partir de información en el dominio de la frecuencia, esto es, utilizando $T_{skin}(\omega)$.

Ahora, para evaluar qué tan bien se representa al modelo IB subglotal mediante el modelo en variables de estado, se propone utilizar como métrica la *raíz del error cuadrático medio* (RMSE, por sus siglas en inglés). Esta métrica, dada una secuencia de referencia $\{y_t\}_{t=1, \dots, N}$

y una secuencia estimada $\{\hat{y}_t\}_{t=1,\dots,N}$ que se quieren comparar, se define como

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}. \quad (3.27)$$

En específico, definimos dos métricas a partir del RMSE que nos servirán para cuantificar el error cometido al intentar replicar la respuesta a impulso h_t y al intentar replicar la respuesta en frecuencia $T_{skin}(\omega)$. Estas métricas están definidas, respectivamente, como

$$\text{RMSE}_t = \sqrt{\frac{1}{(t_1 + 1)} \sum_{t=0}^{t_1} (h_t - \hat{h}_t)^2} \quad (3.28)$$

$$\text{RMSE}_f = \sqrt{\frac{1}{N} \sum_{t=1}^N |T_{skin}(\omega_t) - \hat{T}_{skin}(\omega_t)|^2} \quad (3.29)$$

donde h_t y $T_{skin}(\omega)$ son la respuesta a impulso y respuesta en frecuencia del modelo IB subglotal original y \hat{h}_t y $\hat{T}_{skin}(\omega)$ son la respuesta a impulso y respuesta en frecuencia del modelo en variables de estado. Debemos notar que en el caso del RMSE_t se propone medir el error sólo en el intervalo $[0, t_1]$, donde t_1 es un valor fijo que sirve para considerar sólo la parte central de h_t . Lo anterior es debido a que lo más importante de h_t se concentra en el centro. En este trabajo se considerará $t_1 = 200$.

3.2.1. Reducción de orden

Para comenzar, supongamos que tenemos la respuesta a impulso finita (FIR, por sus siglas en inglés) $\{h_t\}_{t=0,\dots,N}$ del modelo IB subglotal, la cual se relaciona con $T_{skin}(\omega)$ mediante la FFT inversa, esto es, $h_t = \mathcal{F}^{-1}\{T_{skin}(\omega)\}$. Posteriormente, recordando lo que se mostró en la sección 2.4.1 sobre la teoría de realizaciones, es posible utilizar el algoritmo de Zeiger y McEwen, esto es, las ec. (2.28), (2.39), (2.40) y (2.41) para obtener $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s, \mathbf{D}_s)$ a partir de $\{h_t\}_{t=0,\dots,N}$.

Una manera alternativa equivalente para obtener una representación en variables de estado a partir de la respuesta a impulso de un sistema, es utilizando (3.13) en el dominio del tiempo [23]. En específico, transformando dicha ecuación al dominio del tiempo, tenemos que la señal del acelerómetro $y_t^a = \mathcal{F}^{-1}\{\hat{U}_{skin}(\omega)\}$ está dada por

$$y_t^a = h_t * u_t^g = \sum_{m=-\infty}^{\infty} h_m u_{t-m}^g = \sum_{m=0}^N h_m u_{t-m}^g \quad (3.30)$$

donde $*$ representa la convolución en el dominio del tiempo y $u_t^g = \mathcal{F}^{-1}\{U_{sub}(\omega)\}$ es el flujo de aire glotal. Luego, expandiendo la ecuación anterior y realizando ciertas definiciones, tenemos que

$$y_t^a = h_0 u_t^g + h_1 u_{t-1}^g + \dots + h_{N-1} u_{t-N+1}^g + h_N u_{t-N}^g \quad (3.31)$$

$$= \mathbf{C}_s \mathbf{x}_t^s + \mathbf{D}_s u_t^g \quad (3.32)$$

donde se ha definido convenientemente

$$\mathbf{x}_t^s = \begin{bmatrix} u_{t-N}^g & u_{t-N+1}^g & \dots & u_{t-2}^g & u_{t-1}^g \end{bmatrix}^T \in \mathbb{R}^{N \times 1} \quad (3.33)$$

$$\mathbf{C}_s = \begin{bmatrix} h_N & h_{N-1} & \cdots & h_2 & h_1 \end{bmatrix} \in \mathbb{R}^{1 \times N} \quad (3.34)$$

$$\mathbf{D}_s = h_0 \in \mathbb{R}. \quad (3.35)$$

Con la definición anterior de \mathbf{x}_t^s , tenemos que la ecuación de estado (3.25) queda definida por las siguientes matrices de estado y de salida:

$$\mathbf{A}_s = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{N \times N} \quad (3.36)$$

$$\mathbf{B}_s = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^T \in \mathbb{R}^{N \times 1}. \quad (3.37)$$

Así, ya sea utilizando el algoritmo de Zeiger y McEwen o utilizando la construcción mostrada anteriormente, es posible obtener una representación en variables de estado del modelo IB subglotal, entendiéndose que cada método entregará una realización distinta de dicho modelo (esto es, las matrices \mathbf{A}_s , \mathbf{B}_s , \mathbf{C}_s y \mathbf{D}_s serán diferentes pero describirán el mismo modelo).

Una cuestión importante que debemos considerar en este punto, es la causalidad de la respuesta a impulso del modelo IB subglotal. En específico, si observamos la Fig. 3.6, podemos notar que $h_t \neq 0$ para $t < 0$, lo que indica que el sistema es no causal, esto es, el valor de la salida en un instante t es afectada por el valor de la entrada en un instante futuro $t+k$. Lo anterior no es implementable en un modelo en variables de estado pues éstos, intrínsecamente, son modelos causales, por lo que no es posible representar una respuesta a impulso no causal mediante variables de estado. Esto nos condiciona a considerar dos alternativas: la primera, que es considerada en [23], es suponer que en la Fig. 3.6 la respuesta a impulso comienza en $t = 0$, lo que equivale a realizar un corrimiento temporal hacia la derecha de h_t . La segunda alternativa, que es la que se tratará en el presente trabajo de tesis, es sólo considerar la parte causal de h_t , esto es, utilizar \bar{h}_t , la cual aquí definimos como

$$\bar{h}_t = \begin{cases} 0 & \text{si } t < 0 \\ h_t & \text{si } t \geq 0. \end{cases} \quad (3.38)$$

En la Fig. 3.7 se muestra una comparación de las respuestas a impulso y respuestas en frecuencia obtenidas a partir de $T_{skin}(\omega)$ directamente y a partir del modelo en variables de estado de orden 4096 construido a partir de una versión trasladada de h_t . Aquí podemos ver que ambas respuestas a impulso son iguales, aunque la segunda, claramente, está trasladada con respecto a la original. Adicionalmente, la magnitud de ambas respuestas en frecuencia son similares, a diferencia de las fases, donde podemos notar que existe un error que crece linealmente y que puede ser explicado por la traslación temporal de una respuesta a impulso con respecto a la otra.

Por otro lado, en la Fig. 3.8 se muestra la misma comparación, pero ahora el modelo en variables de estado, que es de orden 2048, es construido a partir de la versión causal \bar{h}_t de h_t . Aquí podemos notar que ambas respuestas a impulso son iguales para $t > 0$, pues así fue construida, mientras que las respuestas en frecuencia son muy similares, tanto en magnitud

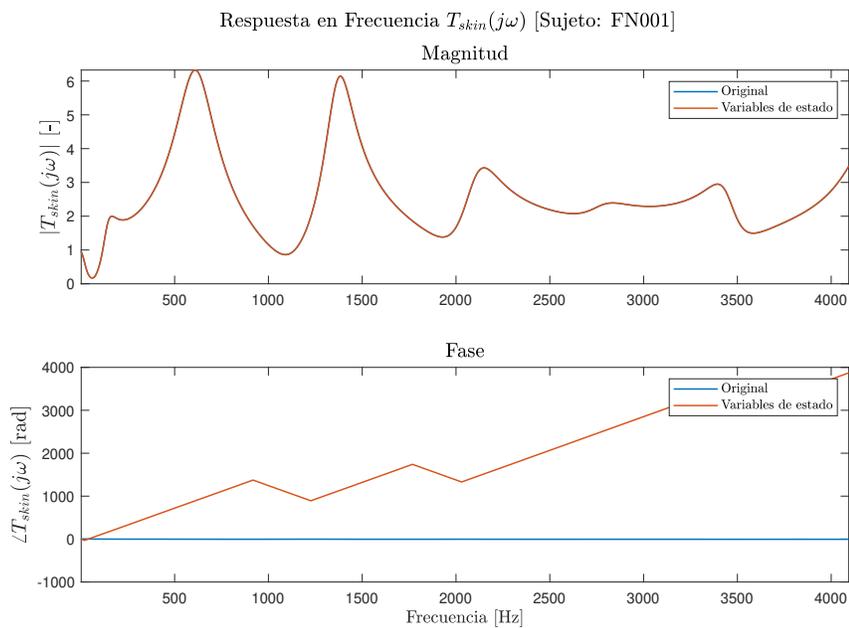
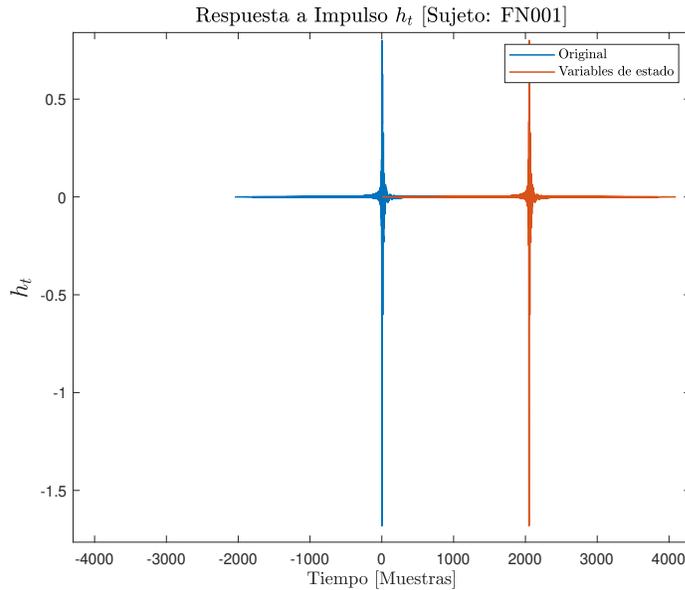
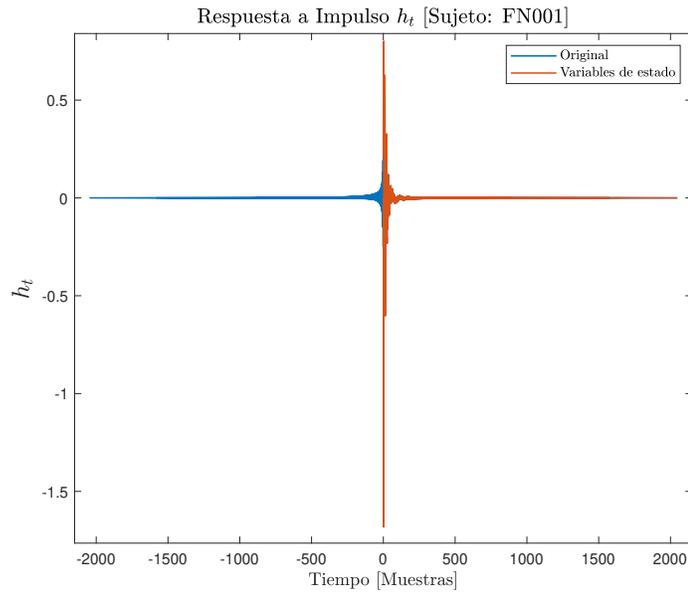
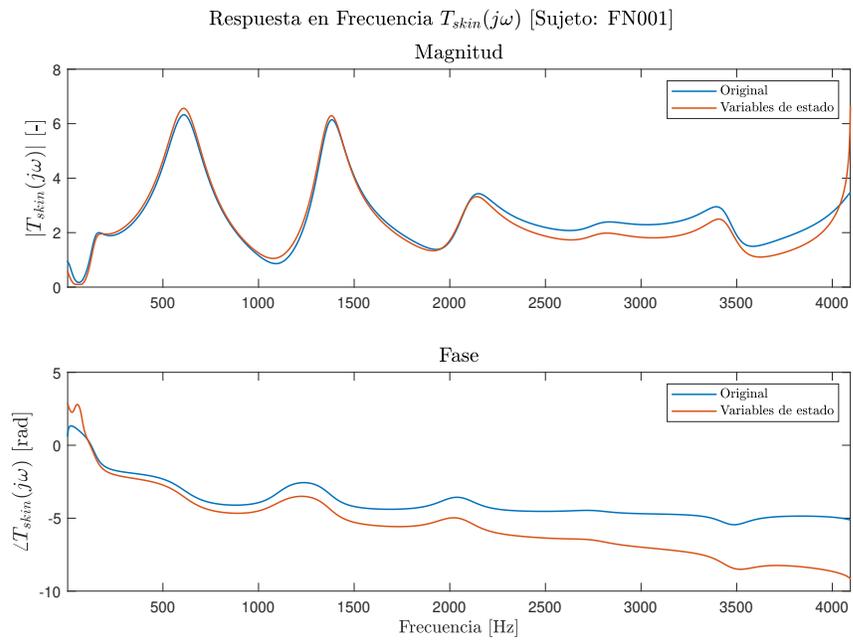


Figura 3.7: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado construido a partir de una versión trasladada de h_t . Los parámetros utilizados fueron los del sujeto FN001.



(a) Respuesta a impulso.



(b) Respuesta en frecuencia.

Figura 3.8: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado construido a partir de la versión causal \bar{h}_t de h_t . Los parámetros utilizados fueron los del sujeto FN001.

como en fase, y donde el error puede ser explicado debido a la pérdida de información que resulta de considerar sólo la parte causal de h_t . Una diferencia adicional entre los modelos en variables de estado obtenidos a partir de h_t trasladada y h_t causal es la diferencia en sus órdenes, siendo el orden del segundo modelo la mitad del orden del primero.

A pesar de que hemos obtenido satisfactoriamente una representación en variables de estado del modelo IB subglotal (ya sea utilizando la versión trasladada o la versión causal de h_t), ésta es de un orden muy alto. En el caso del modelo construido a partir de la versión causal de h_t su orden resultó ser 2048, lo que supondría trabajar con matrices de 2048×2048 generando así un excesivo costo computacional. Por lo anterior, y como uno de los objetivos principales de esta tesis, es necesario encontrar representaciones en variables de estado del modelo IB subglotal cuyos órdenes sean bajos (al menos menores a 100). Así, recordando lo visto en la sección 2.5 sobre reducción de modelos, es posible utilizar el modelo en variables de estado obtenido a partir de la versión causal de h_t como punto de partida y aplicarle una reducción de modelo mediante truncamiento balanceado con el fin de eliminar los estados que no aportan información en la relación entrada-salida.

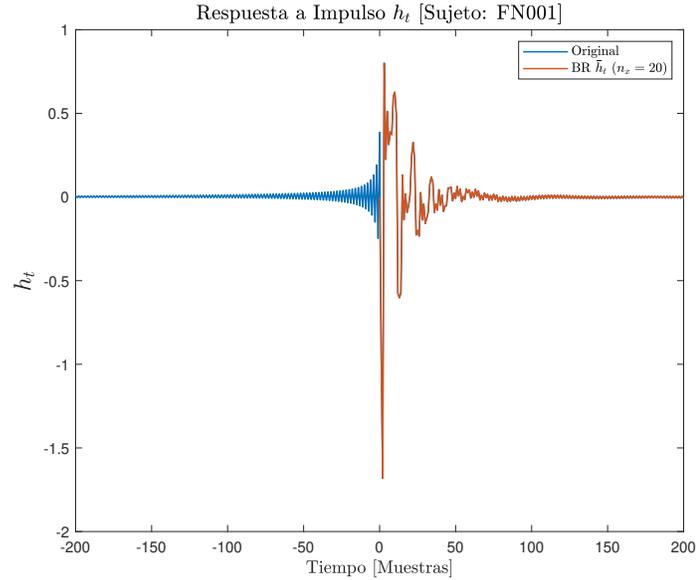
Un ejemplo de aplicar la reducción de modelo mediante truncamiento balanceado al sistema construido a partir de \bar{h}_t se muestra en la Fig. 3.9, donde el sistema reducido es de orden 20. En esta figura, podemos ver que tanto las respuestas a impulso (para $t > 0$) como las respuestas en frecuencias son similares, generando un error parecido al que se observa en la Fig. 3.8, pero con la diferencia de que en este caso el orden del modelo es más de 100 veces menor.

Para analizar un escenario más general, en la Fig. 3.10 se muestran *box plots* del $RMSE_t$ y $RMSE_f$, al comparar las respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con las del modelo obtenido por reducción de orden mediante truncamiento balanceado, para modelos de distintos órdenes y considerando a todos los sujetos de la Tabla 3.1. De la figura podemos notar que en ambas métricas el error parece ser constante entre el orden 15 y 100, lo que nos dice que el modelo en variables de estado construido a partir de \bar{h}_t posee muchos estados que tienen poca relevancia en la relación entrada-salida del sistema y que, por tanto, son posibles de eliminar. En específico, al menos 15 estados son necesarios para mantener una buena aproximación en la relación entrada-salida pues, como podemos ver de la gráfica, cuando se utilizan 5 o 10 estados, el error en la respuesta a impulso aumenta. Como no existe diferencia en cuanto al error al variar entre 15 y 100 estados, es posible elegir órdenes bajos como potenciales representaciones en variables de estado para el modelo IB subglotal.

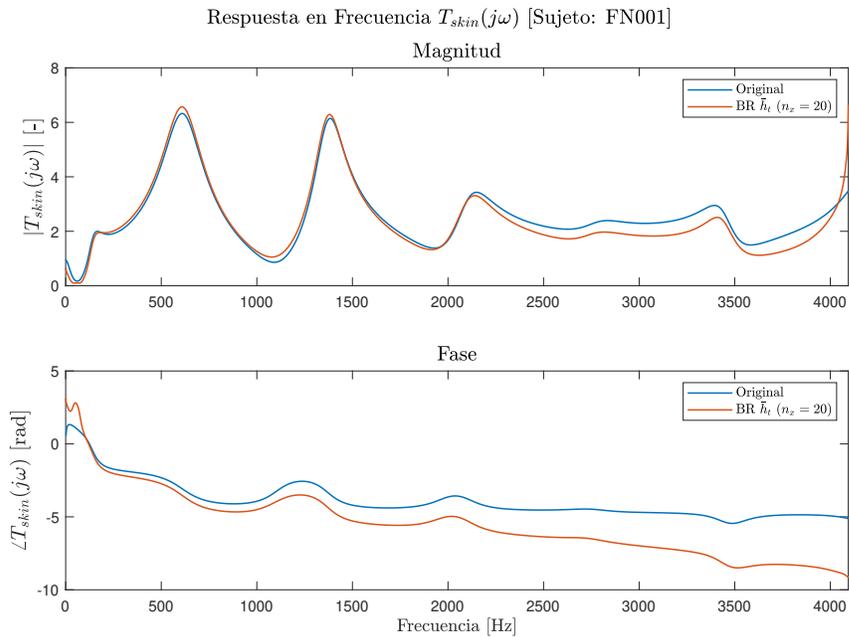
3.2.2. Identificación en el dominio del tiempo

Un enfoque distinto al anterior es el de aplicar técnicas de identificación de sistemas, específicamente las revisadas en las secciones 2.4.2 y 2.4.3. Así, mediante el uso de datos en el dominio del tiempo, representados por la respuesta a impulso h_t del sistema, es posible aplicar métodos de predicción del error en combinación con métodos de subespacios, para obtener una representación en variables de estado para el modelo IB subglotal. En específico, mediante métodos de subespacios, se comienza obteniendo, en un sólo paso, matrices $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s, \mathbf{D}_s)$ para el sistema (3.25)-(3.26). Posteriormente, utilizando estas matrices como inicialización, se aplican métodos de predicción del error.

Ahora bien, en los métodos de identificación de sistemas hay que definir los datos de entrada y salida que utilizará el algoritmo en cuestión. En nuestro caso se utilizará como entrada un impulso unitario, esto es, u_t es un delta de Kronecker δ_t , mientras que como salida



(a) Respuesta a impulso.



(b) Respuesta en frecuencia.

Figura 3.9: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido mediante truncamiento balanceado construido a partir de la respuesta a impulso causal \bar{h}_t . Los parámetros utilizados fueron los del sujeto FN001.

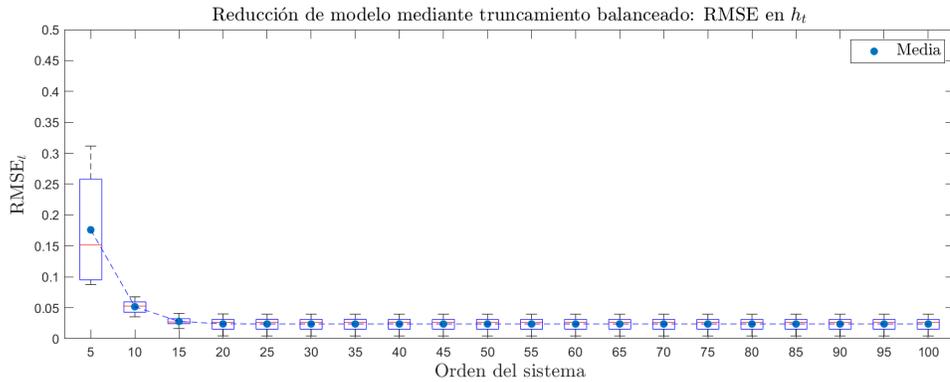
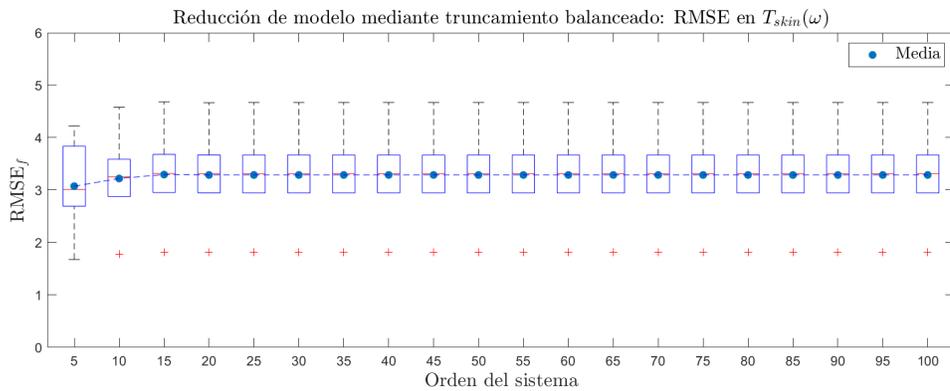
(a) Comparación de respuestas a impulso h_t .(b) Comparación de respuestas en frecuencia $T_{skin}(\omega)$.

Figura 3.10: Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido por reducción de modelo mediante truncamiento balanceado para modelos de distintos órdenes y considerando diferentes sujetos.

se utilizará la respuesta a impulso causal \bar{h}_t del modelo IB subglotal. Un punto importante a considerar está relacionado a la estabilidad del modelo resultante, pues si el sistema es inestable, entonces un impulso unitario resultará en una salida que crecerá indefinidamente. Por lo anterior, al momento de identificar el sistema utilizando PEM, se fuerza a que el modelo resultante sea estable.

En la Fig. 3.11 se muestran las respuestas a impulso y las respuestas en frecuencia del modelo IB subglotal original junto a las del modelo de orden 15 obtenido mediante PEM al utilizar datos en el dominio del tiempo. De la figura podemos notar que en el caso de las respuestas a impulso, la que corresponde al modelo en variables de estado se acerca a la del modelo original, aunque en menor medida que lo obtenido en la sección anterior. Por otro lado, si consideramos las respuestas en frecuencia, notamos que aunque las fases son bastante similares, las magnitudes sólo se parecen en forma (esto es, poseen los *peaks* en las mismas ubicaciones), pues presentan un *error en el escalamiento* por un factor de aproximadamente 2.

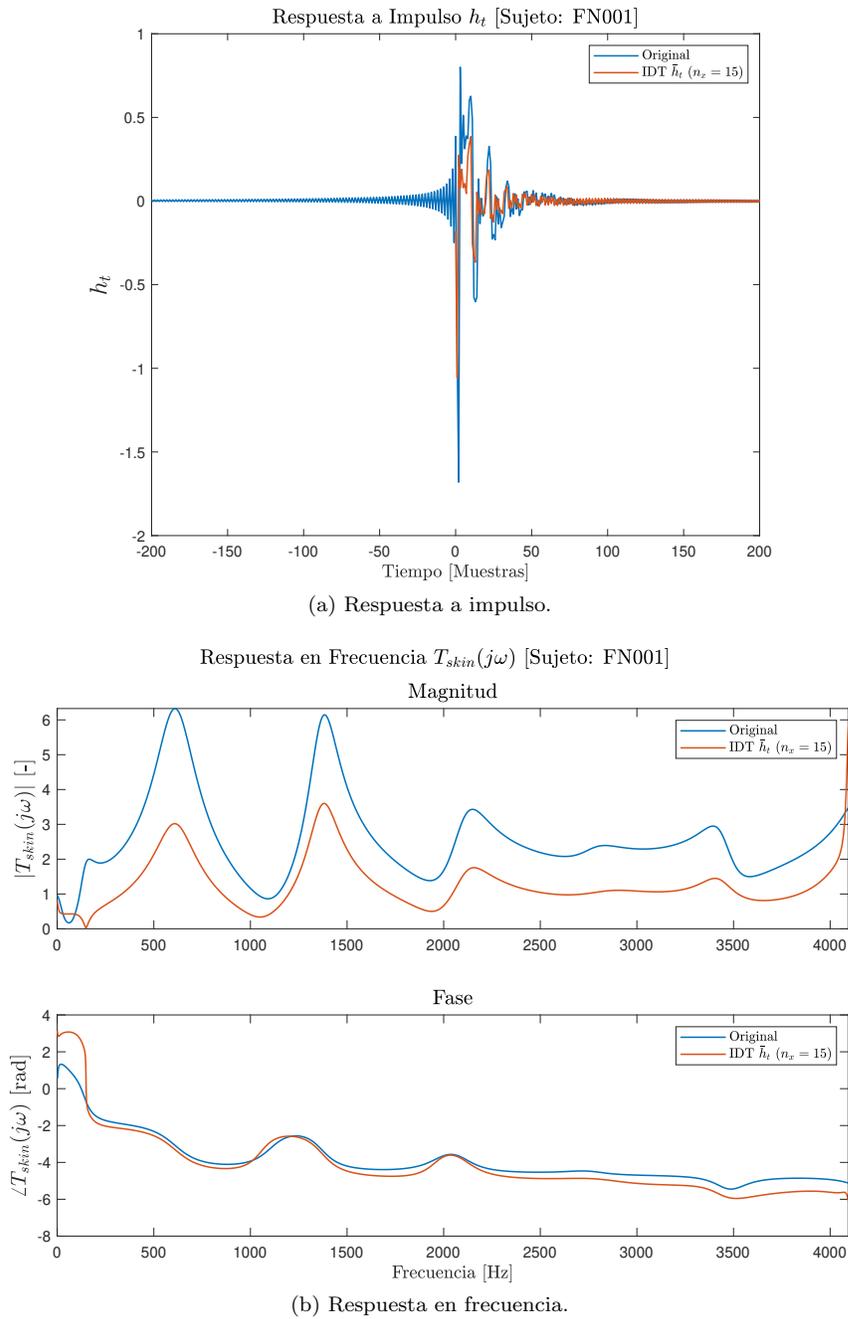


Figura 3.11: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido obtenido mediante identificación en el dominio del tiempo. Los parámetros utilizados fueron los del sujeto FN001.

Al igual que en la sección anterior, en la Fig. 3.12 se muestran *box plots* del $RMSE_t$ y $RMSE_f$ obtenidos al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original con las del modelo obtenido mediante identificación en el dominio del tiempo, para modelos de distintos órdenes y considerando a todos los sujetos de la Tabla 3.1. En este caso, el eje vertical es configurado en base a la sección anterior, de tal manera de facilitar la comparación. Así, podemos ver que en ambos gráficos, y desde el orden 25 en adelante, los valores RMSE son superiores a los límites del eje Y configurados, por lo que no alcanzan a ser mostrados. Sólo es posible observar una pequeña mejora en los órdenes 10 y 15 cuando se considera el RMSE en frecuencia, donde la disminución en el error se obtiene debido a que, como vimos en la Fig. 3.11, este método entrega modelos cuyas fases se acercan más a la del modelo original. A pesar de lo anterior, si ahora consideramos el RMSE en las respuestas a impulso, podemos notar que existe el doble del error que en la sección anterior. Así, es posible concluir que la identificación en el dominio del tiempo entrega modelos en variables de estado cuyas respuestas a impulso y en frecuencia se acercan en menor medida a las del modelo original en comparación con los modelos obtenidos en la sección anterior.

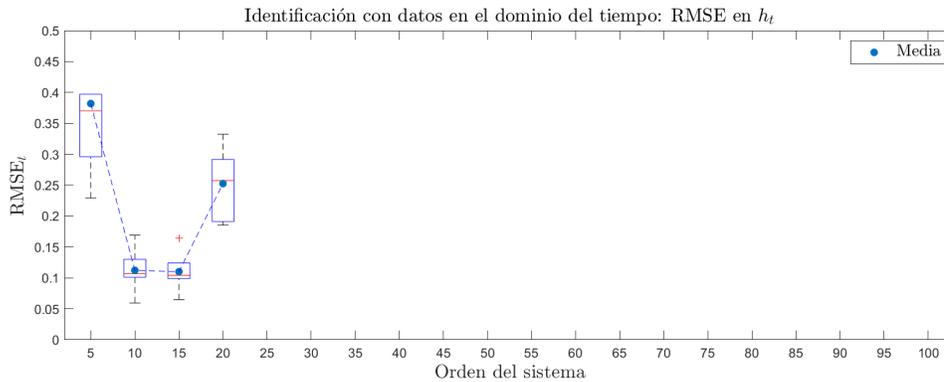
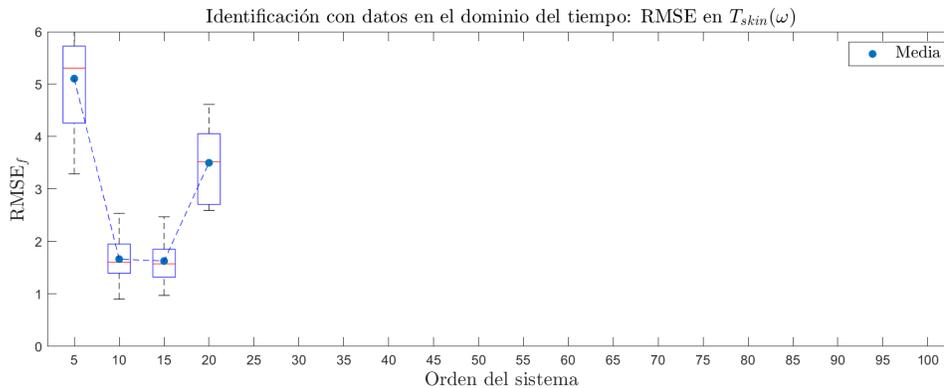
(a) Comparación de respuestas a impulso h_t .(b) Comparación de respuestas en frecuencia $T_{skin}(\omega)$.

Figura 3.12: Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido mediante identificación en el dominio del tiempo para modelos de distintos órdenes y considerando diferentes sujetos.

3.2.3. Identificación en el dominio de la frecuencia

Ahora bien, las técnicas de identificación revisadas en las secciones 2.4.2 y 2.4.3 también son posibles de aplicar utilizando datos en el dominio de la frecuencia. En específico, es posible utilizar la respuesta en frecuencia del modelo IB subglotal, esto es, $T_{skin}(\omega)$, de tal manera de obtener un modelo en variables de estado cuya respuesta en frecuencia sea lo más próxima a ella. Un punto importante a considerar aquí es que, para que las matrices $(\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s, \mathbf{D}_s)$ resulten con coeficientes reales, es necesario que se cumpla que $T_{skin}(0) \in \mathbb{R}$ (pues, un modelo con coeficientes reales debe tener fase 0 a frecuencia continua). Por lo anterior y debido a que usualmente $T_{skin}(0) \in \mathbb{C}$, redefinimos el valor de $T_{skin}(\omega)$ en $\omega = 0$ como $T_{skin}(0) := |T_{skin}(0)|$. Además, igual que en la sección anterior, se fuerza a que modelo resultante sea estable.

A continuación, en la Fig. 3.13 se muestran las respuestas a impulso y en frecuencia del modelo IB subglotal original junto a las del modelo de orden 20 obtenido mediante PEM al utilizar datos en el dominio de la frecuencia. Al revisar las respuestas en frecuencia podemos notar claramente que la correspondiente al modelo en variables de estado es bastante similar a la del modelo original, tanto en magnitud como en fase. Incluso, es posible advertir que las respuestas en frecuencia exhibidas aquí son más parecidas que en el caso de utilizar reducción de modelo mediante truncamiento balanceado. Si ahora consideramos las respuestas a impulso, también podemos corroborar que, para $t \geq 0$, ambas respuestas son técnicamente idénticas.

Para validar de manera más general lo mencionado anteriormente, en la Fig. 3.14 se muestran *box plots* del $RMSE_t$ y $RMSE_f$ al comparar las respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con las del modelo obtenido mediante identificación en el dominio de la frecuencia, para distintos órdenes de modelo y considerando a todos los sujetos de la Tabla 3.1. En dichas figuras, cuyos ejes verticales se han configurado igual que en las secciones anteriores, podemos ver que en la mayoría de los modelos obtenidos los errores son pequeños. De hecho, si comparamos el valor de los errores obtenidos aquí con aquellos exhibidos en la sección 3.2.1, podemos notar que ambos tipos de errores ($RMSE_t$ y $RMSE_f$) son menores, donde la diferencia más notoria se observa en el valor de $RMSE_f$, cuyo valor aquí es más de 7 veces menor, en promedio. Lo anterior tiene sentido debido a que el modelo obtenido mediante identificación en el dominio de la frecuencia lo que hace es precisamente disminuir el error en la respuesta en frecuencia, esto es, $RMSE_f$. Una observación importante, es que la respuesta a impulso también resulta ser idéntica al considerar la parte causal, lo que permite corroborar que el modelo obtenido es una buena aproximación del modelo IB subglotal. Lo anterior subraya una ventaja de la identificación en el dominio de la frecuencia por sobre los métodos anteriores, y es que al utilizar la respuesta en frecuencia como dato para la identificación, no es necesario *eliminar* información como en los casos anteriores (donde se tuvo que descartar arbitrariamente el valor de h_t para $t < 0$), pues la propiedad de causalidad es agregada implícitamente por la estructura de nuestro modelo en variables de estado. Además, el modelo IB subglotal es construido precisamente en el dominio de la frecuencia, por lo que obtener un modelo en variables de estado que tenga su misma respuesta en frecuencia implica conseguir un modelo más consecuente con el modelamiento inicial.

En conclusión, al comparar los resultados obtenidos por los tres métodos propuestos en la tarea de obtener una representación en variables de estado de bajo orden para el modelo IB subglotal, es posible decir que el método que mejor resultados obtiene es el de identificación utilizando datos en el dominio de la frecuencia. Debido a que varios modelos de bajo orden

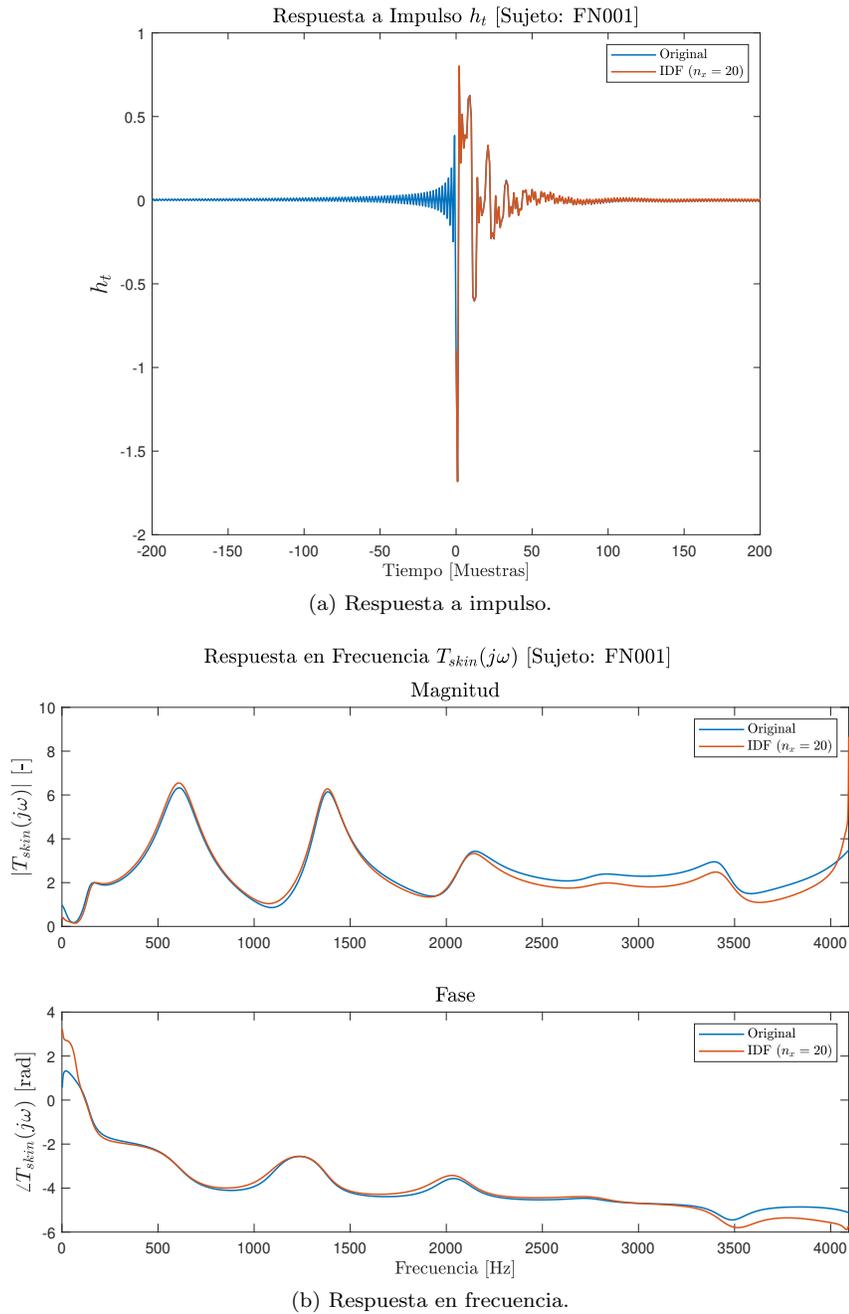


Figura 3.13: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo de orden reducido obtenido mediante identificación en el dominio de la frecuencia. Los parámetros utilizados fueron los del sujeto FN001.

entregan valores similares para $RMSE_t$ y $RMSE_f$, aquí se eligen los 5 órdenes que mejores resultados dieron en promedio. En específico, en los capítulos posteriores se considerarán los modelos de orden 15, 20, 25, 30 y 35 obtenidos mediante identificación en el dominio de la frecuencia.

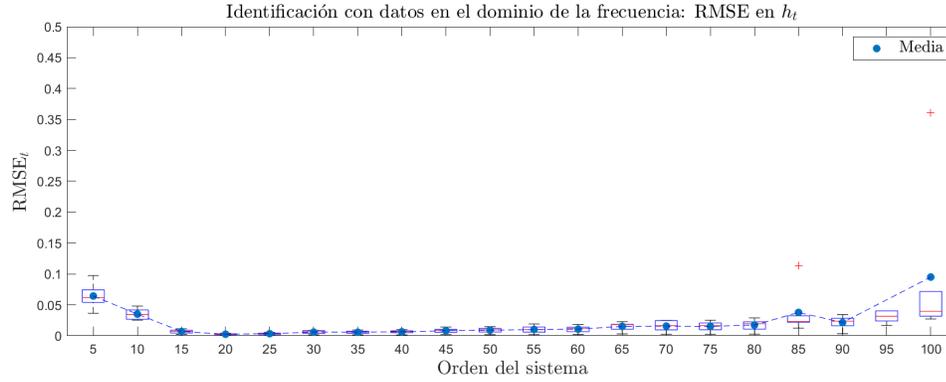
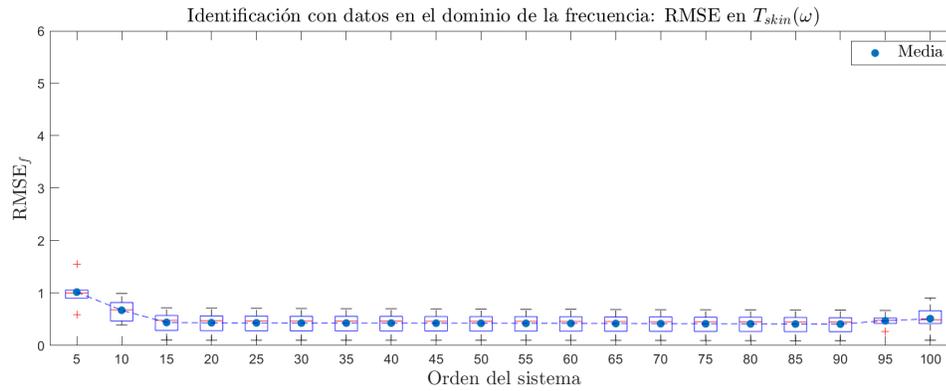
(a) Comparación de respuestas a impulso h_t .(b) Comparación de respuestas en frecuencia $T_{skin}(\omega)$.

Figura 3.14: Raíz del error cuadrático medio al comparar las respuestas a impulso y en frecuencia del modelo IB subglotal original y su modelo equivalente en variables de estado obtenido mediante identificación en el dominio de la frecuencia para modelos de distintos órdenes y considerando diferentes sujetos.

FILTRAJE INVERSO MEDIANTE SUAVIZADOR DE KALMAN

Una vez obtenida una representación en variables de estado del modelo basado en impedancia subglotal, es posible utilizarla para obtener estimaciones del flujo de aire glotal a partir de las mediciones de la aceleración que existe sobre la piel del cuello. En este capítulo se trata el problema del filtraje inverso basado en impedancia subglotal (IBIF), el cual se refiere al problema de obtener la señal del flujo de aire glotal, a partir de la aceleración sobre la piel del cuello, haciendo uso del modelo IB subglotal. En específico, se comienza el capítulo explicando qué es el filtraje inverso y cómo se aplica para obtener el flujo de aire glotal a partir del flujo de aire oral. Luego, se presenta el enfoque referido como IBIF, el cual tiene como objetivo obtener estimaciones del flujo de aire glotal a través de la inversión del modelo IB subglotal en el dominio de la frecuencia. Posteriormente, se expone un enfoque Bayesiano, el cual busca estimar la señal de interés mediante un filtro de Kalman aplicado a un modelo IB subglotal en variables de estado de alto orden. Finalmente, se estudia el enfoque propuesto en la presente tesis, el cual busca mejorar los enfoques anteriores mediante el uso de un suavizador de Kalman aplicado al modelo IB subglotal, el cual es representado mediante un modelo en variables de estado de bajo orden.

4.1. Filtraje inverso

Como se ha señalado en capítulos anteriores, la teoría fuente-filtro sitúa al flujo de aire que pasa por la glotis como la fuente sonora que, al pasar por el tracto vocal y la radiación de los labios, genera la señal del habla (Fig. 4.1). Matemáticamente, esto es equivalente a decir que la señal que representa al flujo de aire glotal pasa a través de filtros, los cuales le dan la forma para dar como resultado la señal del habla.

Ahora bien, el flujo de aire glotal es una señal que proporciona información valiosa, pues, en esta teoría, es la fuente principal de la señal del habla y además, como se ha dicho a lo largo de este trabajo, es una señal que posee características que resultan útiles en el diagnóstico de la hiperfunción vocal. Sin embargo, obtenerla de forma directa resulta una tarea complicada, pues lo anterior supondría, al menos, una intervención quirúrgica complicada con el fin de

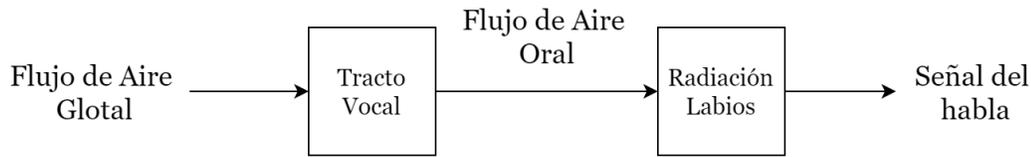


Figura 4.1: Modelo de la teoría fuente-filtro [61].

posicionar algún dispositivo en la glotis, el cual sea capaz de adquirir mediciones. Como lo anterior no es factible, es necesario recurrir a la medición de otras señales, las cuales nos permitan obtener el flujo de aire glotal de manera indirecta.

Una primera opción para realizar lo anterior sería utilizar la señal del habla (esto es, la presión radiada por nuestra boca y labios), la cual puede ser obtenida por un micrófono, para posteriormente aplicar una técnica llamada *filtraje inverso*, la cual es utilizada para estimar la fuente de excitación en la glotis (en este caso, el flujo de aire glotal) a partir de la remoción de las cargas acústicas de una señal de salida (en este caso, la señal del habla) [22]. En específico, las funciones de transferencia del tracto vocal y de la radiación de los labios son invertidas, lo que resulta en una nueva función de transferencia, la cual representa un filtro inverso y que tendrá un polo y un cero por cada cero y polo, respectivamente, de las funciones de transferencia originales (Fig. 4.2).

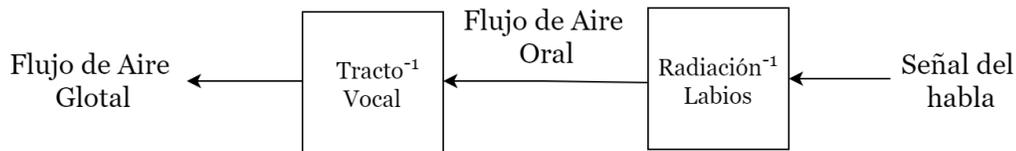


Figura 4.2: Filtraje inverso para la obtención del flujo de aire glotal.

En la Fig. 4.3 se muestra un esquema del filtraje inverso a partir de la señal del habla para obtener el flujo de aire glotal (*Glottal V. V. Waveform*, en la figura). En esta figura, podemos ver que la señal que entra al filtro inverso (*Inverse Filter*, en la figura) es la obtenida por un micrófono puesto a cierta distancia de los labios. Además, podemos observar que la respuesta en frecuencia del tracto vocal junto a la radiación de los labios (cuyo gráfico es mostrado a la izquierda) posee ciertos *peaks* de resonancias (llamados *formantes*), los cuales equivalen a las resonancias acústicas del tracto vocal. Dicha respuesta en frecuencia también posee un cero en el origen, el cual es introducido por la radiación de los labios, la cual es representada por el término $j\omega$. Adicionalmente, se muestra la respuesta en frecuencia del filtro inverso, el cual posee un cero y un polo por cada polo y cero, respectivamente, de las funciones de transferencia originales.

Las principales desventajas del filtraje inverso a partir de la señal del habla provienen del polo en el origen (equivalente a un integrador en el dominio del tiempo) que debe ser incluido para cancelar el cero en el origen agregado por la radiación de los labios [14, 62]. Lo anterior se debe a que dicho polo incrementa la respuesta a bajas frecuencias, por lo que este método es muy sensible a ruido de baja frecuencia. Además, debido al integrador y a que no se conocen las condiciones iniciales, existe una pérdida del nivel de flujo a frecuencia cero

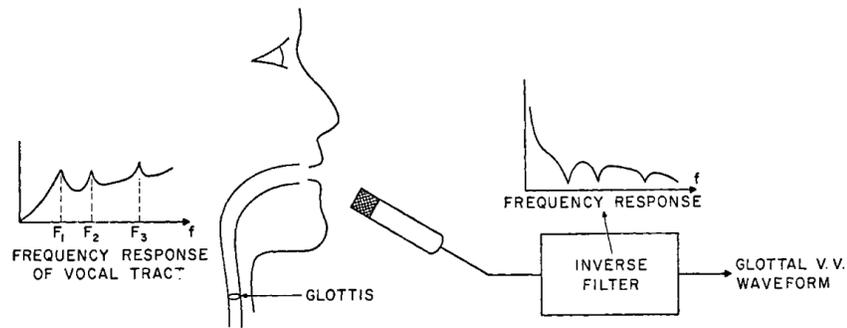


Figura 4.3: Filtrado inverso a partir de la señal del habla [62]. El gráfico de la respuesta en frecuencia del tracto vocal incluye la radiación de los labios.

(en otras palabras, el nivel DC). Adicionalmente, la calibración de la amplitud se hace difícil debido a la dependencia de factores tales como la distancia y orientación entre el micrófono y los labios. Finalmente, esta forma de filtrado inverso no toma en cuenta la nasalización¹, el cual es un proceso importante en la producción de ciertos sonidos [14, 62].

Como solución a dichas desventajas, en [14] se propuso una nueva técnica de filtrado inverso, donde ahora la señal de salida utilizada para obtener el flujo de aire glotal es el flujo de aire que sale *directamente* del tracto vocal, sin considerar el efecto de los labios, el cual es conocido como flujo de aire oral (Fig. 4.2). En la Fig. 4.4 se muestra un esquema del filtrado inverso a partir del flujo de aire oral, el cual es obtenido a partir de un neumatógrafo con ventilación circunferencial, más conocido como máscara de Rothenberg (Fig. 4.5). Adicionalmente, en la Fig. 4.6 se muestra un ejemplo de las formas de onda del flujo de aire oral y del flujo de aire glotal cuando un sujeto de prueba pronuncia la vocal /a/. El flujo de aire oral fue obtenido mediante una máscara de Rothenberg, mientras que el flujo de aire glotal fue obtenido a partir de la técnica de filtrado inverso aplicado al flujo de aire oral, donde ahora sólo es necesario invertir el filtro relativo al tracto vocal.

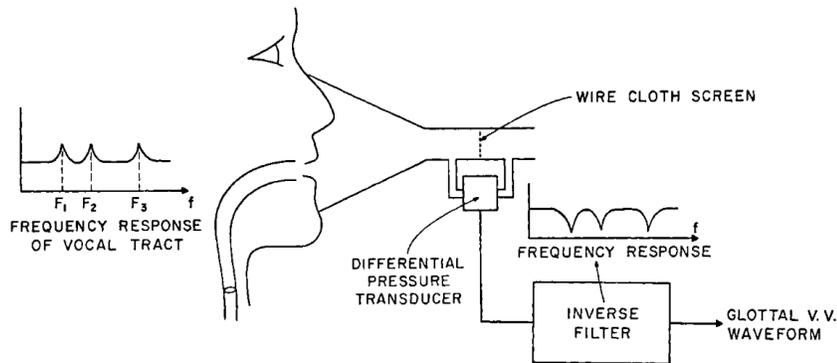


Figura 4.4: Filtrado inverso a partir del flujo de aire oral [14, 62].

¹La nasalización es la producción de un sonido mientras el velo del paladar está abajo, de tal manera que algo de aire escapa a través de la nariz.

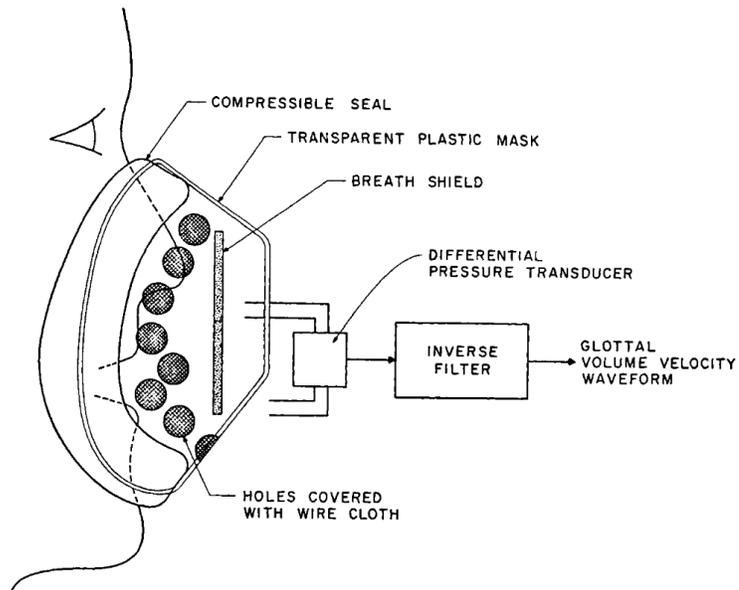


Figura 4.5: Neumotacógrafo con ventilación circunferencial, más conocido como máscara de Rothenberg [14].

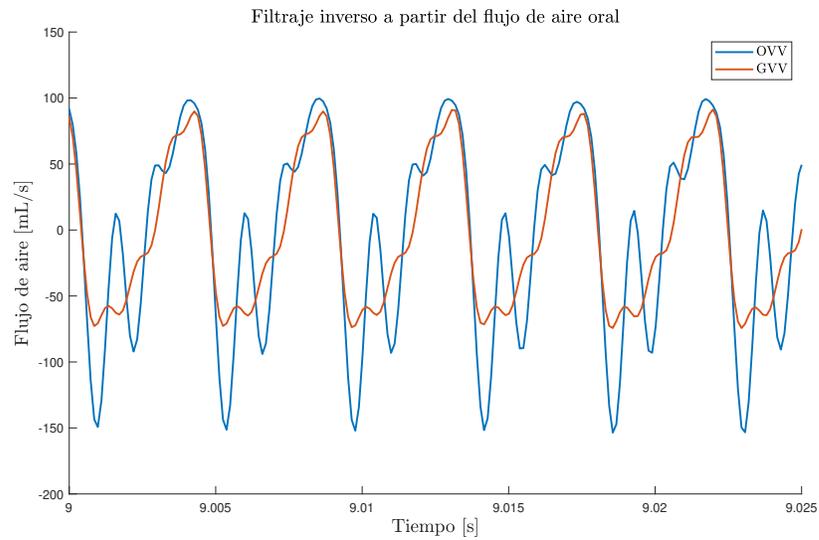


Figura 4.6: Flujo de aire glotal (GVV, en rojo) obtenido mediante el filtraje inverso del flujo de aire oral (OVV, en azul). Éste último fue obtenido mediante una máscara de Rothenberg puesta sobre un sujeto de prueba que pronunciaba la vocal /a/.

La principal ventaja de utilizar este método, es que el modelo inverso no necesita un polo en el origen (esto es, un integrador) pues la radiación de los labios no influye en el flujo de aire oral. Lo anterior soluciona el problema de la sensibilidad al ruido de baja frecuencia y del desconocimiento del nivel de flujo a frecuencia cero. Además, la calibración de la amplitud tampoco supone un problema, pues la máscara mide el flujo de aire total de la boca y la nariz, el cual es igual al flujo de aire glotal a bajas frecuencias, por lo que la calibración puede ser fácilmente realizada mediante un flujo de aire constante [14].

Debido a las amplias ventajas mencionadas anteriormente y a la precisión presentada, el filtraje inverso a partir del flujo de aire oral [14] es una técnica que entrega estimaciones confiables del flujo de aire glotal. De hecho, debido a la precisión de dicha estimación, ésta es utilizada como señal de referencia en estudios que buscan estimar el flujo de aire glotal (por ejemplo, en [22, 23]).

A pesar de sus múltiples ventajas y precisión, esta técnica debe aplicarse en un ambiente de laboratorio, con profesionales que dirijan la obtención del flujo de aire oral mediante la manipulación de la máscara de Rothenberg. Es por lo anterior que en la siguientes secciones se presenta un método alternativo, el cual busca obtener estimaciones del flujo de aire glotal de manera ambulatoria y no invasiva, para compararlo con el obtenido a partir del flujo oral que se considerará como señal de referencia o señal *ground truth*.

4.2. Filtraje inverso basado en impedancia subglotal

En el capítulo 3 se mostró el modelo basado en impedancia subglotal, el cual relaciona al flujo de aire glotal con la aceleración sobre la piel del cuello en el dominio de la frecuencia (Fig. 4.7). En específico, este modelo considera como entrada al flujo de aire glotal y como salida la señal del acelerómetro. Como podemos notar, esta es una situación parecida a la mencionada en la sección anterior, ya que el flujo de aire glotal nuevamente entra a un filtro (en este caso, el modelo basado en impedancia subglotal), el cual le da la forma para dar como resultado la aceleración sobre la piel del cuello (la cual puede ser medida mediante un acelerómetro). Así, basados en la idea de filtraje inverso, en [22] se propuso la obtención del flujo de aire glotal a partir del filtrado inverso de la señal del acelerómetro. A dicha técnica se le denominó *filtraje inverso basado en impedancia subglotal* (IBIF subglotal, por sus siglas en inglés), pues hace uso del modelo basado en impedancia subglotal para obtener el filtro inverso, tal como se muestra en el esquema de la Fig. 4.8.

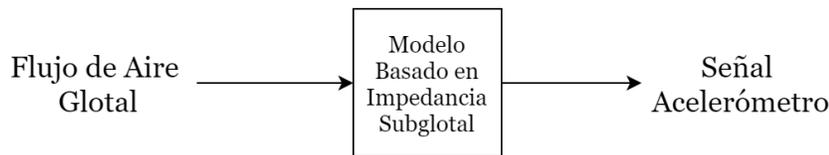


Figura 4.7: Modelo basado en impedancia subglotal [22].

Ahora bien, para entender cómo se lleva a cabo el enfoque de IBIF subglotal, debemos recordar algunas expresiones del capítulo 3. En específico, en la sección 3.1 se obtuvo la expresión (3.13), la cual relaciona, en el dominio de la frecuencia, al modelo basado en impedancia subglotal $T_{skin}(\omega)$, con la señal del acelerómetro $\dot{U}_{skin}(\omega)$ y el flujo de aire

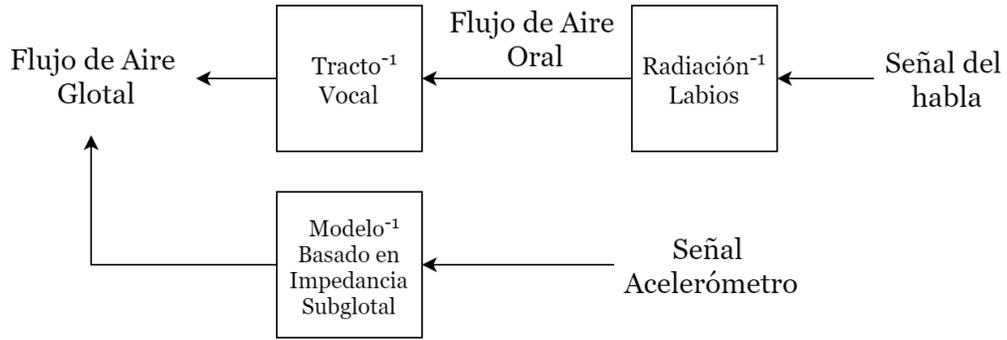


Figura 4.8: Filtraje inverso basado en impedancia subglotal [22]. A modo de comparación se muestra el filtraje inverso mediante otras señales posibles.

glotal $U_{sub}(\omega)$ de la siguiente manera:

$$T_{skin}(\omega) = \frac{\dot{U}_{skin}(\omega)}{U_{sub}(\omega)}. \quad (4.1)$$

Así, el flujo de aire glotal $U_{sub}(\omega)$ puede ser obtenido mediante la inversión de la respuesta en frecuencia $T_{skin}(\omega)$, es decir,

$$U_{sub}(\omega) = \frac{\dot{U}_{skin}(\omega)}{T_{skin}(\omega)}. \quad (4.2)$$

Notemos que de momento todas las variables y funciones son de tiempo continuo, sin embargo la señal del acelerómetro es una señal de tiempo discreto, pues éste obtiene mediciones de la aceleración cada ciertos intervalos de tiempo. Por lo anterior, es necesario realizar el filtraje inverso en tiempo discreto, lo cual se lleva a cabo utilizando la FFT junto a la IFFT [63]. En específico, y tal como se dijo en el capítulo anterior, $T_{skin}(\omega)$ es obtenida para un conjunto discreto de frecuencias $\omega_k = 2\pi f_s k/N$ con $k = 0, 1, \dots, N-1$ y donde f_s es la frecuencia de muestreo y N es el número de puntos de la FFT. Con lo anterior obtenemos la transformada discreta de Fourier, $T_{skin}[k] = T_{skin}(\omega_k)$, que representa el muestreo de $T_{skin}(\omega)$ en el dominio de la frecuencia. Finalmente, para obtener una estimación del flujo de aire glotal de tiempo discreto \hat{u}_t^g , se debe calcular la FFT de la señal del acelerómetro y_t^a , esto es, $\dot{U}_{skin}[k] = \mathcal{F}\{y_t^a\}$, para luego multiplicarla por $1/T_{skin}[k]$ y luego calcularle la IFFT al producto resultante, esto es:

$$\hat{u}_t^g = \mathcal{F}^{-1} \left\{ \frac{\dot{U}_{skin}[k]}{T_{skin}[k]} \right\}. \quad (4.3)$$

Una forma alternativa y equivalente para obtener el flujo de aire glotal mediante IBIF subglotal, es realizando lo anterior en el dominio del tiempo. Específicamente, si obtenemos la respuesta a impulso \tilde{h}_t (equivalente a la IFFT) de la inversa del modelo IB subglotal, esto es, $\tilde{h}_t = \mathcal{F}^{-1}\{1/T_{skin}[k]\}$, entonces la ec. (4.3) es equivalente a la siguiente convolución:

$$\hat{u}_t^g = \tilde{h}_t * y_t^a = \sum_{m=-\infty}^{\infty} \tilde{h}_m y_{t-m}^a. \quad (4.4)$$

En la Fig. 4.9-a se muestra (en rojo) el flujo de aire glotal estimado obtenido mediante el enfoque de IBIF subglotal, cuando el sujeto de prueba pronuncia la vocal /a/. Además, como señal de referencia (en azul) se utiliza el flujo de aire glotal obtenido mediante el filtrado inverso del flujo de aire oral. Por otro lado, en la Fig. 4.9-b se muestra (en rojo) la derivada del flujo de aire glotal obtenido mediante IBIF subglotal y (en azul) la derivada de la señal de referencia. En ambos gráficos, podemos ver que la señal estimada y la de referencia se parecen bastante en cuanto a forma, amplitud y fase, a pesar de ser obtenidas a partir de dos técnicas y señales totalmente distintas. Más específicamente, y desde un punto de vista cuantitativo, el RMSE² cometido por las estimaciones obtenidas a partir de IBIF es de 8.3881, lo que, al normalizarlo por $(u_{\text{máx}}^g - u_{\text{mín}}^g)$, equivale a un RMSE normalizado (NRMSE, por sus siglas en inglés) de 5.1020 %. Lo anterior sugiere que el enfoque IBIF subglotal es un método equivalente para obtener estimaciones del flujo de aire glotal de una forma ambulatoria y no invasiva, convirtiéndose así en un método potencialmente apropiado para, por ejemplo, el monitoreo y la evaluación de la hiperfunción vocal durante las tareas diarias de los pacientes o para enfoques de biorretroalimentación que buscan facilitar y mejorar la terapia conductual de la voz [22]. En efecto, estudios realizados posteriormente [17, 21] han mostrado que el esquema de IBIF subglotal entrega estimaciones del flujo de aire glotal cuyas medidas aerodinámicas son comparables a las obtenidas mediante el filtrado inverso del flujo de aire oral, en términos de diferenciar una función vocal normal y una función vocal hiperfuncional (incluidos tanto PVH y NPVH).

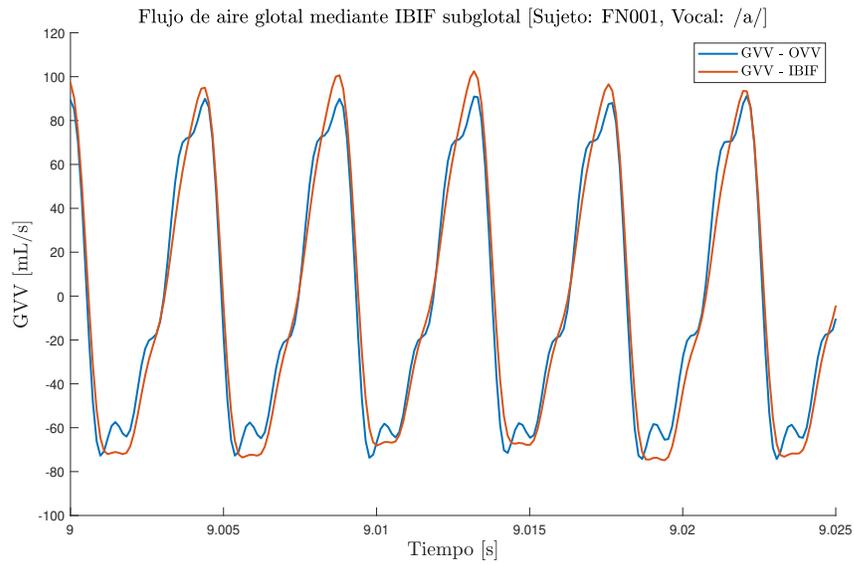
El enfoque de IBIF subglotal requiere de los parámetros Q que definen al modelo IB subglotal de un sujeto, los cuales son obtenidos mediante un esquema de calibración (ver sección 3.1) que es dependiente de la señal de referencia (y, por tanto, de la vocal pronunciada) utilizada para llevarlo a cabo. Por tanto, las estimaciones del flujo de aire glotal se ven afectadas por la calibración del modelo IB subglotal. Para ejemplificar lo anterior, en la Fig. 4.10 se muestran estimaciones del flujo de aire glotal junto a su derivada a partir del esquema IBIF subglotal (en rojo), además de la señal de referencia obtenida a partir de OVV (en azul), cuando las señales filtradas son obtenidas a partir de la vocal /i/ y el modelo IB subglotal es calibrado con la vocal /a/. En ambas figuras podemos ver que, aunque las estimaciones se acercan a la señal de referencia, esta vez la similitud no es tanta como la obtenida para el caso de la vocal /a/, lo cual puede ser corroborado al calcular el RMSE cometido al estimar GVV, el cual entrega un valor de 10.9623 (equivalente a un NRMSE de 6.3121 %) lo que es un 30.6887 % superior al caso de la vocal /a/.

Por otro lado, la segunda necesidad está ligada a la falta de alguna métrica que permita cuantificar la calidad de las estimaciones del flujo de aire glotal, la cual sería una fuente de información valiosa considerando que las estimaciones son obtenidas en un contexto ambulatorio, donde no existe una señal de referencia.

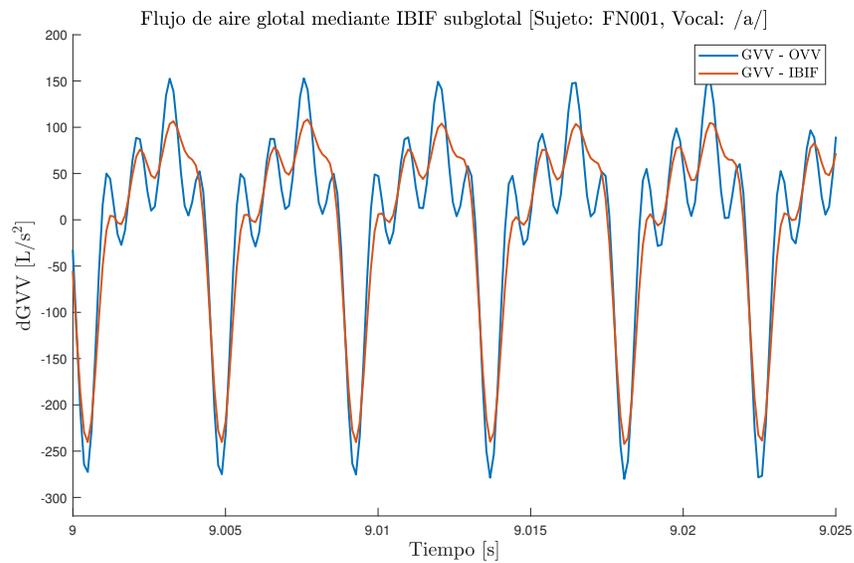
4.3. Filtraje inverso mediante filtraje de Kalman de alto orden

Un enfoque que parece apropiado para resolver los problemas descritos en la sección anterior, es el de filtraje y suavizamiento Bayesiano (descritos en la sección 2.6) pues, debido a su naturaleza, permite estimar señales de una forma probabilística, entregando así una medida de la incertidumbre en las estimaciones (en específico, la matriz de covarianza del error de las estimaciones), además de *adaptar* las estimaciones con respecto al modelo y la señal del acelerómetro, lo que supondría una mejora en las estimaciones aún cuando la vocal cambie.

²Esta y otras métricas serán definidas apropiadamente en el capítulo 5.

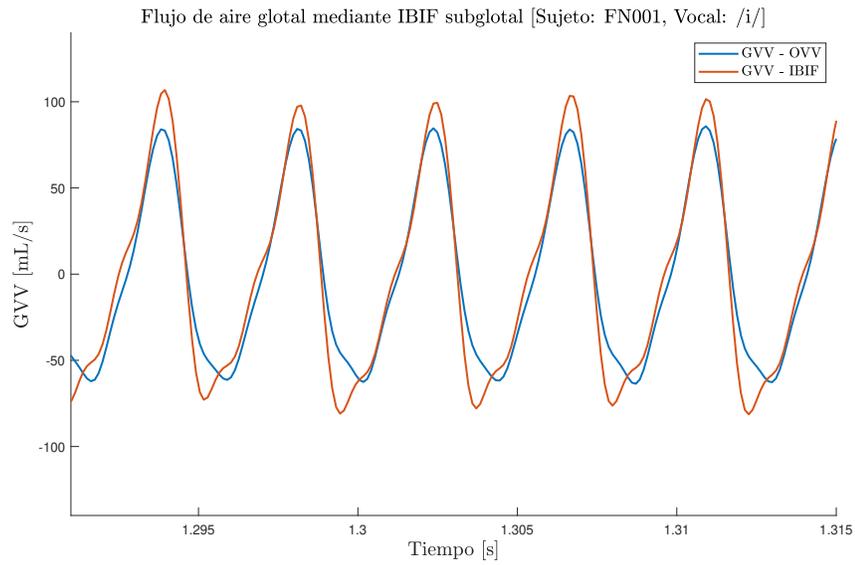


(a) Flujo de aire glotal.

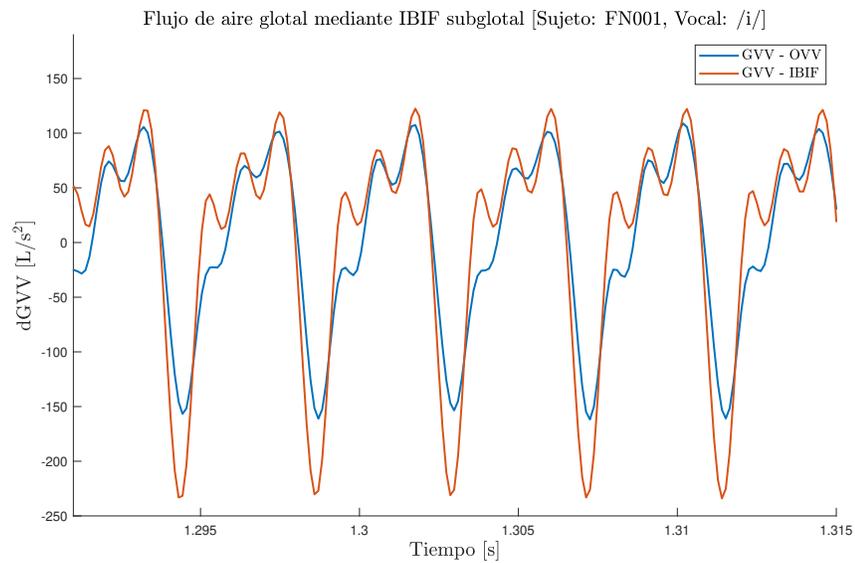


(b) Derivada del flujo de aire glotal.

Figura 4.9: Flujo de aire glotal (GVV, en rojo) junto a su derivada (dGVV, en rojo) obtenidos mediante IBIF subglotal. A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV). La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /a/.



(a) Flujo de aire glotal.



(b) Derivada del flujo de aire glotal.

Figura 4.10: Flujo de aire glotal (GVV, en rojo) junto a su derivada (dGVV, en rojo) obtenidos mediante IBIF subglotal a partir de la vocal /i/. A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV). La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /i/.

Basado en lo anterior, en [23] se propone un esquema Bayesiano, donde ahora el filtraje inverso se realiza mediante un filtro de Kalman. En específico, en [23] se comienza proponiendo una representación en variables de estado del modelo IB subglotal, tal como el obtenido al comienzo de la sección 3.2, pero con la diferencia de que ahora el modelo es estocástico. Para comprender lo anterior, recordemos que una representación en variables de estado del modelo IB subglotal está dada por

$$\mathbf{x}_{t+1}^s = \mathbf{A}_s \mathbf{x}_t^s + \mathbf{B}_s u_t^g \quad (4.5)$$

$$y_t^a = \mathbf{C}_s \mathbf{x}_t^s + \mathbf{D}_s u_t^g \quad (4.6)$$

donde u_t^g es el flujo de aire glotal, y_t^a es la señal del acelerómetro y \mathbf{x}_t^s es el vector de estados del sistema. Además, dicho vector de estados está definido como

$$\mathbf{x}_t^s = \begin{bmatrix} u_{t-N}^g & u_{t-N+1}^g & \cdots & u_{t-2}^g & u_{t-1}^g \end{bmatrix}^\top \in \mathbb{R}^{N \times 1} \quad (4.7)$$

mientras que las matrices están dadas por

$$\mathbf{A}_s = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{N \times N} \quad (4.8)$$

$$\mathbf{B}_s = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^\top \in \mathbb{R}^{N \times 1} \quad (4.9)$$

$$\mathbf{C}_s = \begin{bmatrix} h_N & h_{N-1} & \cdots & h_2 & h_1 \end{bmatrix} \in \mathbb{R}^{1 \times N} \quad (4.10)$$

$$\mathbf{D}_s = h_0 \in \mathbb{R}. \quad (4.11)$$

De las definiciones anteriores, podemos ver primeramente que, por la forma en que se llevó a cabo la construcción del modelo en variables de estado, la señal a estimar u_t^g se encuentra implícitamente en el estado \mathbf{x}_t^s . En efecto, el estado del sistema está compuesto por versiones de u_t^g desplazadas temporalmente, siendo la primera componente del estado igual al valor de u_t^g hace N muestras y la última componente igual al valor de u_t^g hace una muestra. Por lo anterior y dado que el filtro de Kalman estima el estado de un sistema, entonces es posible aplicarlo en este caso para estimar implícitamente el flujo de aire glotal u_t^g a partir de la estimación del estado \mathbf{x}_t^s . Aquí es importante notar que debido a la composición del estado, cuando se estima \mathbf{x}_t^s utilizando información hasta el instante t (esto es, realizando filtraje), se están estimando implícitamente versiones pasadas de u_t^g . En específico, al hacer filtraje y utilizar información sobre y_t^a hasta el instante t , se están estimando $u_{t-N}^g, \dots, u_{t-1}^g$, siendo, por ejemplo, \hat{u}_{t-N}^g una estimación suavizada de u_{t-N}^g . Por lo anterior, en [23] se menciona que se está implementando un suavizador de Kalman, a pesar de que se utiliza el algoritmo y las ecuaciones del filtro de Kalman.

Ahora bien, el sistema descrito por (4.5)-(4.6) es un sistema determinístico cuya entrada *determinística* u_t^g es precisamente la señal que se quiere estimar. Para ejecutar el filtro de Kalman sería necesario conocer dicha entrada para todo t , pero justamente es la señal de flujo de aire glotal que se desea obtener. Por lo anterior, en [23] se supone que la entrada

u_t^g es una señal estocástica w_t , específicamente ruido blanco Gaussiano, esto es, $u_t^g = w_t$ con $w_t \sim \mathcal{N}_w(0, \sigma_w^2)$. Al mismo tiempo, debido a que las mediciones del acelerómetro no son exactas, esta incertidumbre se agrega en el modelo al suponer que la señal de salida y_t^a contiene una componente de ruido aditivo v_t , la cual se supone ruido blanco Gaussiano, esto es, $v_t \sim \mathcal{N}_v(0, \sigma_v^2)$. Con esto, tenemos que el sistema determinístico se convierte en el siguiente sistema estocástico sin entradas³:

$$\mathbf{x}_{t+1}^s = \mathbf{A}_s \mathbf{x}_t^s + \mathbf{w}_t \quad (4.12)$$

$$y_t^a = \mathbf{C}_s \mathbf{x}_t^s + v_t \quad (4.13)$$

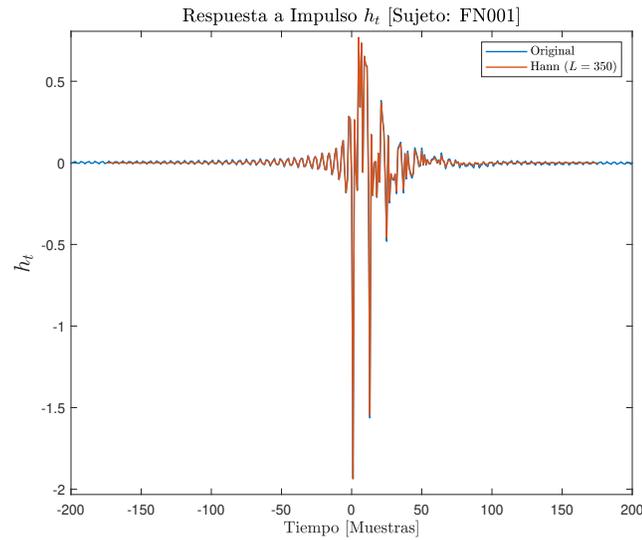
donde v_t es el ruido de salida y $\mathbf{w}_t = \mathbf{B}_s w_t$ es el ruido de proceso, con $\mathbf{w}_t \sim \mathcal{N}_w(\mathbf{0}, \mathbf{Q})$ y

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_w^2 \end{bmatrix} \in \mathbb{R}^{N \times N}. \quad (4.14)$$

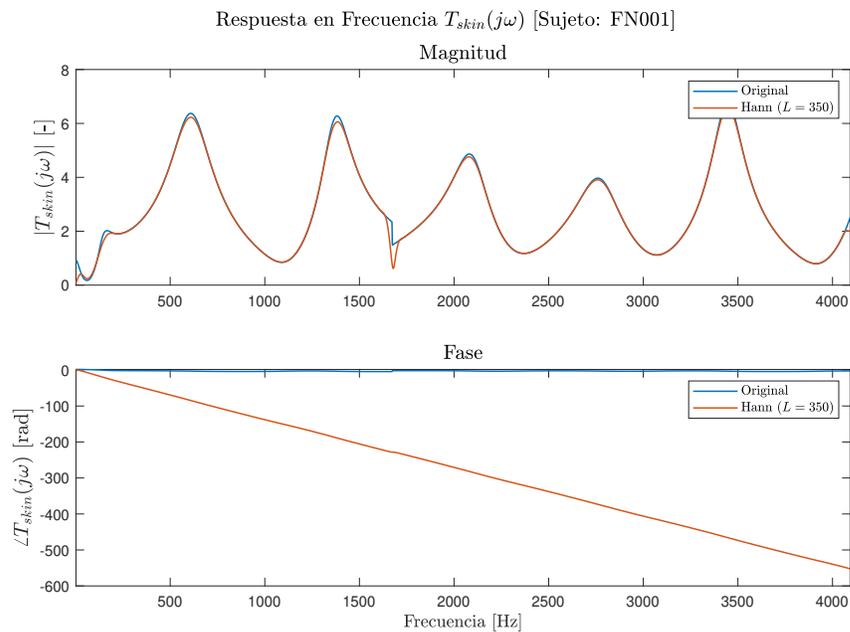
Con lo anterior, al sistema estocástico descrito por las ec. (4.12)-(4.13) se le puede aplicar las ecuaciones del filtro de Kalman. A pesar de lo anterior, existe una limitante en este esquema, la cual está relacionada con el orden del modelo anterior. Como podemos ver de las definiciones de las matrices que describen este sistema, tenemos que $\mathbf{A}_s \in \mathbb{R}^{N \times N}$ y $\mathbf{C}_s \in \mathbb{R}^{1 \times N}$, donde N es el largo de la respuesta a impulso utilizada para construirlo. Al mismo tiempo, si recordamos que la relación entre la respuesta a impulso h_t y la respuesta en frecuencia $T_{skin}[k] = T_{skin}(\omega_k)$ está dada por la IFFT inversa $h_t = \mathcal{F}^{-1}\{T_{skin}[k]\}$ entonces podemos notar que N es el número de puntos utilizados para muestrear $T_{skin}(\omega)$. En efecto, entre mayor sea el valor de N , mayor será la precisión al discretizar $T_{skin}(\omega)$, pero también mayor será el orden del sistema y con ello la complejidad computacional (tanto en tiempo de cómputo como en memoria); al contrario, entre menor sea el valor de N , menor será el orden del sistema y la complejidad computacional, pero también menor será la precisión lograda al discretizar $T_{skin}(\omega)$. Lo anterior sugiere que este método requiere de un balance en la elección de N de tal manera de no comprometer la precisión de $T_{skin}[k]$ ni la complejidad computacional.

Para resolver lo anterior, y debido a que h_t concentra su *energía* principalmente en el centro (ver Fig. 3.6), en [23] se propone obtener $T_{skin}[k]$ (y consecuentemente h_t) con un valor alto de N (en general $N > 550$), para luego aplicar una ventana de Hann a h_t y así obtener una versión truncada de h_t [63]. Luego de hacer un pequeño estudio en cuanto a la pérdida de precisión utilizando distintos largos de ventanas L , en [23] se concluye que $L = 350$ es un largo apropiado (aunque conservador) que permite obtener una representación en variables de estado cuyo orden es computacionalmente manejable además de tener una respuesta a impulso cuyo RMSE es menor a 0.1 cuando se compara con la respuesta a impulso original. En la Fig. 4.11-a se muestra un ejemplo de la respuesta a impulso truncada en el centro con una ventana de Hann de largo $L = 350$, mientras que la Fig. 4.11-b muestra la respuesta en frecuencia del modelo en variables de estado construido a partir de dicha respuesta a impulso. En ambos casos podemos ver que las gráficas coinciden, con excepción de la fase de

³En este caso se ha supuesto arbitrariamente que no existe paso directo, esto es, $\mathbf{D}_s = 0$.



(a) Respuesta a impulso.



(b) Respuesta en frecuencia.

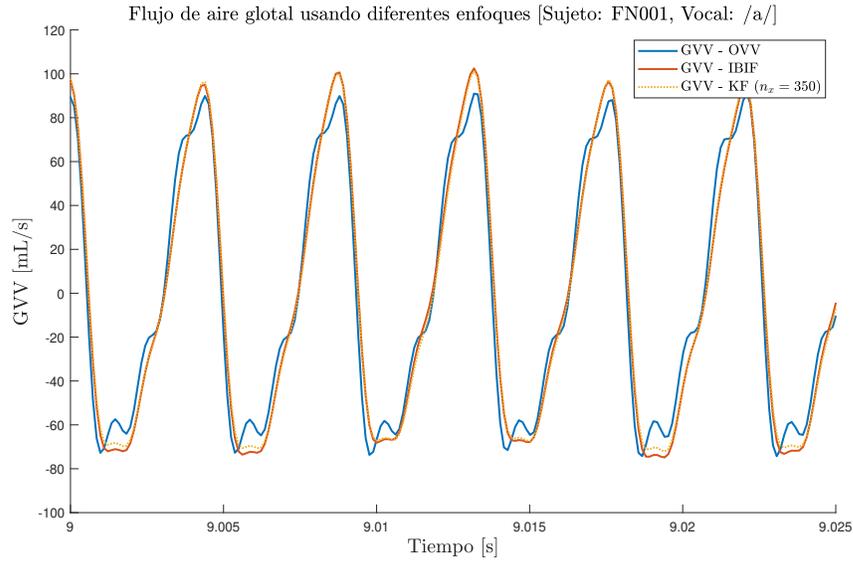
Figura 4.11: Comparación de respuestas a impulso y respuestas en frecuencia del modelo IB subglotal original con la del modelo en variables de estado obtenido a partir de la respuesta a impulso truncada en el centro con una ventana de Hann de largo $L = 350$. La respuesta a impulso del sistema en variables de estado se ha desplazado hacia la izquierda para efectos de comparación. Los parámetros usados fueron los del sujeto FN001.

la respuesta en frecuencia debido a que, como se vio en el capítulo 3, la respuesta a impulso del modelo en variables de estado, intrínsecamente causal, es una versión retrasada en $N/2$ muestras de la respuesta a impulso original, la cual es no causal.

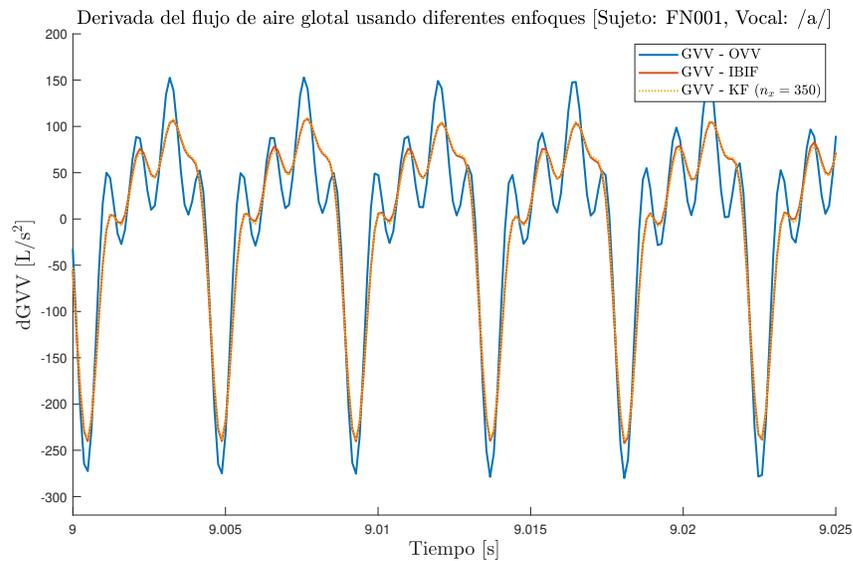
Los parámetros de diseño del filtro de Kalman son las varianzas de los ruidos w_t y v_t . En [23] se resuelve dicha elección mediante la búsqueda del par (σ_w^2, σ_v^2) que entregue el menor RMSE al estimar el flujo de aire glotal. En específico, en dicho trabajo se encuentra que una amplia gama de combinaciones entregan el RMSE mínimo, eligiendo finalmente $\sigma_w^2 = 100$ y $\sigma_v^2 = 1$ a lo largo del escrito. En la Fig. 4.12 se muestra (en amarillo) la estimación del flujo de aire glotal junto a su derivada utilizando el esquema de filtraje de Kalman, junto a la señal de referencia (en azul), obtenida a partir del filtraje inverso de OVV, y la estimación realizada por el esquema IBIF subglotal (en rojo). En ambos casos podemos ver que las estimaciones obtenidas mediante filtraje de Kalman siguen muy de cerca a las estimaciones obtenidas mediante IBIF subglotal, lo cual es natural considerando que ambos esquemas utilizan el mismo modelo para obtener las estimaciones. Además, podemos observar que las estimaciones obtenidas por ambos esquemas se acercan bastante a la señal de referencia, con pequeñas diferencias que no permiten distinguir cuál estimación es mejor. Para lo anterior, podemos obtener el RMSE cometido por ambas estimaciones, donde el esquema IBIF genera un RMSE de 8.3881 (equivalente a un NRMSE de 5.1020 %), mientras que el esquema de filtraje de Kalman genera un RMSE de 8.1494 (equivalente a un NRMSE de 4.9481 %). En este caso, podemos confirmar que el filtro de Kalman entrega mejores estimaciones del flujo de aire glotal que el esquema IBIF, incluso aunque los valores de σ_w^2 y σ_v^2 escogidos no necesariamente son los mejores para este caso, pues fueron obtenidos para otros sujetos y, por ende, para un escenario distinto. Así, si por ejemplo eligiéramos $\sigma_w^2 = 100$ y $\sigma_v^2 = 10$, entonces el RMSE generado por el filtro de Kalman disminuye a un 7.8404 (equivalente a un NRMSE de 4.7491 %), entregando de esta forma mejores estimaciones que en el caso anterior. Este último hecho supone una pequeña ventaja para el filtro de Kalman con respecto a IBIF, en el sentido de que el primero posee parámetros adicionales (estos son, σ_w^2 y σ_v^2) que, al ser ajustados, permitirían *mejorar* las estimaciones.

Ahora bien, si consideramos el caso de estimar el flujo de aire glotal a partir de señales obtenidas con la vocal /i/, las estimaciones también mejoran con respecto al esquema de IBIF subglotal. Por ejemplo, en la Fig. 4.13 se muestran las estimaciones del flujo de aire glotal junto a su derivada obtenidas mediante IBIF subglotal (en rojo) y filtraje de Kalman (en amarillo), acompañadas de la señal de referencia (en azul), a partir del filtraje inverso de las señales relacionadas a la vocal /i/. En este caso, podemos ver que las estimaciones del filtro de Kalman se acercan a la señal de referencia levemente más que las obtenidas mediante IBIF, lo cual puede ser cuantificado por el RMSE, cuyo valor en el caso de IBIF es de 10.9623 (equivalente a un NRMSE de 6.3121 %), mientras que en el caso del filtro de Kalman es de 10.7145 (equivalente a un NRMSE de 6.1668 %).

A pesar de los positivos resultados obtenidos en [23] para las estimaciones del flujo de aire glotal mediante el uso del filtro de Kalman, se debe considerar el costo computacional tanto en tiempo de cómputo como en memoria. Solo por dar un ejemplo, al procesar 107 383 muestras obtenidas por el acelerómetro (equivalentes a 13.1083 segundos de adquisición de datos), el esquema IBIF demora un tiempo de 0.0264 segundos en procesar los datos y obtener estimaciones del flujo de aire glotal, en comparación con los 240.6437 segundos utilizados por el filtro de Kalman. Además, éste último, al ser una solución iterativa, requiere mantener en memoria matrices y vectores de orden 350, lo que podría no ser implementable

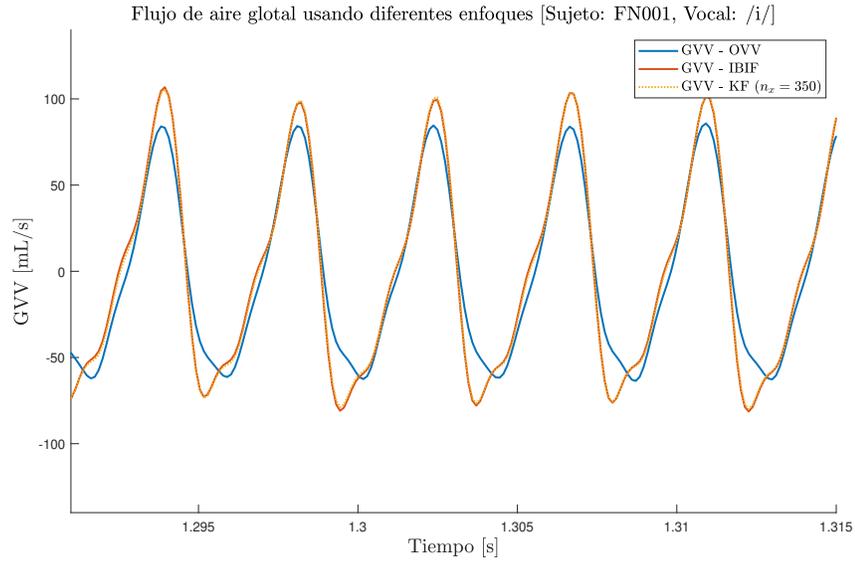


(a) Flujo de aire glotal.

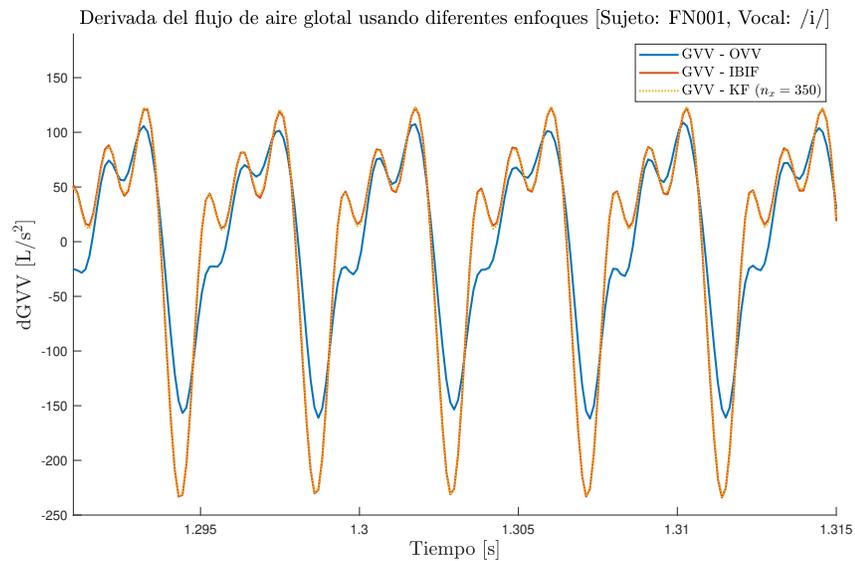


(b) Derivada del flujo de aire glotal.

Figura 4.12: Flujo de aire glotal (GVV, en amarillo) junto a su derivada (dGVV, en amarillo) obtenidos mediante filtraje de Kalman (KF). A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV) y las estimaciones del flujo de aire glotal (GVV, en rojo) junto a su derivada (dGVV, en rojo) a partir de IBIF subglotal. La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /a/.



(a) Flujo de aire glotal.



(b) Derivada del flujo de aire glotal.

Figura 4.13: Flujo de aire glotal (GVV, en amarillo) junto a su derivada (dGVV, en amarillo) obtenidos mediante filtraje de Kalman (KF) a partir de la vocal /i/. A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV) y las estimaciones del flujo de aire glotal (GVV, en rojo) junto a su derivada (dGVV, en rojo) a partir de IBIF subglotal. La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /i/.

en un dispositivo cuyo fin es ser *wearable*⁴.

Un último comentario sobre el trabajo presentado en [23] es que en él se estudian modelos del flujo de aire glotal de tal manera de modelar u_t^g ya no como un simple ruido blanco Gaussiano, sino como ruido coloreado, el cual es incluido en las ec. (4.12)-(4.13). Sin embargo, el realizar lo anterior, además de aumentar el orden del sistema, no entregó mejoras sustanciales en los resultados, por lo que posteriormente estos modelos no fueron considerados.

Es por las debilidades que presenta el esquema de filtraje de Kalman de alto orden, que en la siguiente sección se busca dar una mejora a éstas, con el fin de obtener mejores estimaciones del flujo de aire glotal en comparación con IBIF, al mismo tiempo de mantener un bajo costo computacional, de tal manera de que el cambio de esquema sea beneficioso.

4.4. Filtraje inverso mediante suavizamiento de Kalman de bajo orden

En esta sección se presenta la principal contribución de esta tesis. Como se ha dicho en la sección anterior, el esquema propuesto en [23] se basa en un modelo en variables de estado de alto orden. Para resolver lo anterior, en el capítulo 3 se estudió la obtención de una representación en variables de estado para el modelo basado en impedancia subglotal sujeto a que dicha representación fuera de bajo orden (al menos de orden inferior a 100, esto es, más de 3 veces menor que en [23]). En efecto, utilizando los datos originales del modelo IB subglotal, esto es, la respuesta en frecuencia $T_{skin}(\omega)$, fue posible encontrar modelos en variables de estado de bajo orden (entre 15 y 35) que representaban de muy buena manera el modelo original, ya sea al comparar las respuestas en frecuencia como al comparar las respuestas a impulso. Lo anterior, nos abre la posibilidad de aplicar un esquema Bayesiano, como el de suavizamiento de Kalman, para la resolución del problema a un costo computacional menor.

Un posible problema de utilizar el modelo reducido es la pérdida del significado de los estados durante el proceso de la identificación del sistema. Esto es, aunque nuevamente tenemos un modelo en variables de estado de la forma

$$\mathbf{x}_{t+1}^s = \mathbf{A}_s \mathbf{x}_t^s + \mathbf{B}_s u_t^g \quad (4.15)$$

$$y_t^a = \mathbf{C}_s \mathbf{x}_t^s + \mathbf{D}_s u_t^g, \quad (4.16)$$

esta vez las matrices que definen el modelo no poseen una estructura pre-establecida, dando como resultado que las componentes del vector de estados \mathbf{x}_t^s no posean un significado como en el caso mostrado en la sección anterior (donde el estado contenía versiones corridas temporalmente de la señal a estimar).

Para resolver el problema anterior, es necesario introducir, de alguna manera, la entrada u_t^g que se desea estimar, en el estado \mathbf{x}_t^s . Es claro de que al realizar lo anterior, el vector de estado aumentará de dimensión y con ello la dimensión de las matrices que definen al sistema. En las siguientes subsecciones, se presentan distintas maneras de resolver el problema anterior.

4.4.1. Modelando la entrada como ruido blanco (WT)

Un enfoque bastante simple para introducir la entrada en el estado, el cual no supone un aumento considerable de dimensionalidad, es el sugerido en [64], donde la idea fundamental

⁴Término utilizado para referirnos a dispositivos que se llevan puestos como accesorios.

se describe a continuación. En primera instancia, re-definimos el vector de estados de la siguiente manera

$$\tilde{\mathbf{x}}_t^s = \begin{bmatrix} \mathbf{x}_t^s & u_t^g \end{bmatrix}^\top. \quad (4.17)$$

Con lo anterior, tenemos que el sistema descrito por (4.15)-(4.16), puede ser re-escrito como

$$\tilde{\mathbf{x}}_{t+1}^s = \tilde{\mathbf{A}}_s \tilde{\mathbf{x}}_t^s + \tilde{\mathbf{B}}_s \tilde{u}_t^g \quad (4.18)$$

$$y_t^a = \tilde{\mathbf{C}}_s \tilde{\mathbf{x}}_t^s \quad (4.19)$$

con

$$\tilde{\mathbf{A}}_s = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s \\ \mathbf{0} & 0 \end{bmatrix} \in \mathbb{R}^{(n_x+1) \times (n_x+1)} \quad (4.20)$$

$$\tilde{\mathbf{B}}_s = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^\top \in \mathbb{R}^{(n_x+1) \times 1} \quad (4.21)$$

$$\tilde{\mathbf{C}}_s = \begin{bmatrix} \mathbf{C}_s & \mathbf{D}_s \end{bmatrix} \in \mathbb{R}^{1 \times (n_x+1)} \quad (4.22)$$

donde n_x es el orden del sistema original, esto es, $\mathbf{A}_s \in \mathbb{R}^{n_x \times n_x}$ y donde \tilde{u}_t^g es el flujo de aire glotal adelantado en una muestra con respecto al flujo de aire glotal a estimar u_t^g . Esto último puede ser comprobado con la última ecuación de (4.18), donde tenemos que $\tilde{x}_{t+1, n_x+1}^s = u_{t+1}^g = \tilde{u}_t^g$. En efecto, la representación alternativa (4.18)-(4.19) posee como estado adicional una versión atrasada de la entrada, por lo que este estado es el único afectado directamente por la entrada, para posteriormente encargarse de excitar a los estados originales.

Ahora bien, como podemos darnos cuenta de la ec. (4.18), aún existe el término determinístico \tilde{u}_t^g , el cual deberíamos conocer para poder ejecutar un suavizador de Kalman (y lo que no tendría mucha lógica porque es esencialmente la señal que queremos estimar). En [64] se trata este problema al suponer que la entrada \tilde{u}_t^g , al ser desconocida, puede ser modelada como una señal estocástica, específicamente como ruido blanco Gaussiano con una varianza alta. Así, podemos suponer que $\tilde{u}_t^g = w_t$ con $w_t \sim \mathcal{N}(0, \sigma_w^2)$, para luego tener como ruido de proceso a $\mathbf{w}_t = \tilde{\mathbf{B}}_s w_t$ con $\mathbf{w}_t \sim \mathcal{N}_w(\mathbf{0}, \mathbf{Q})$ y con

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_w^2 \end{bmatrix} \in \mathbb{R}^{(n_x+1) \times (n_x+1)}. \quad (4.23)$$

Al mismo tiempo, debido a que las mediciones del acelerómetro no son exactas, esta incertidumbre se agrega en el modelo al suponer que y_t^a contiene una componente de ruido aditivo v_t , la cual se supone ruido blanco Gaussiano, esto es $v_t \sim \mathcal{N}_v(0, \sigma_v^2)$. Con lo anterior, tenemos que el sistema determinístico (4.18)-(4.19) se convierte en el siguiente sistema estocástico sin entradas:

$$\tilde{\mathbf{x}}_{t+1}^s = \tilde{\mathbf{A}}_s \tilde{\mathbf{x}}_t^s + \mathbf{w}_t \quad (4.24)$$

$$y_t^a = \tilde{\mathbf{C}}_s \tilde{\mathbf{x}}_t^s + v_t. \quad (4.25)$$

Con lo anterior, es posible utilizar un suavizador de Kalman, para así estimar el vector de estados $\tilde{\mathbf{x}}_t^s$, el cual contiene el flujo de aire glotal u_t^g como su última componente. En la Fig. 4.14 es muestra el flujo de aire glotal (junto a su derivada) obtenido mediante los métodos anteriormente descritos (IBIF subglotal y filtraje de Kalman de alto orden) junto a las estimaciones obtenidas mediante el suavizador de Kalman al modelar el flujo de aire glotal como ruido blanco Gaussiano y considerando una representación en variables de estado de orden 26 con varianzas iguales a $\sigma_w^2 = 1000$ y $\sigma_v^2 = 100$. En la figura podemos ver que las estimaciones son bastante buenas y, de hecho, se acercan más a la señal de referencia que los otros métodos. Lo anterior puede ser confirmado si consideramos el RMSE cometido al estimar el flujo de aire glotal mediante el suavizador de Kalman, el cual tiene un valor de 8.0694 (y un NRMSE de 4.9486 %), los cuales son menores a los obtenidos con los métodos anteriores: 8.3881 (equivalente a un NRMSE de 5.1020 %) para IBIF y 8.1494 (equivalente a un NRMSE de 4.9481 %) para el filtro de Kalman.

4.4.2. Modelando la entrada como Random Walk (RW)

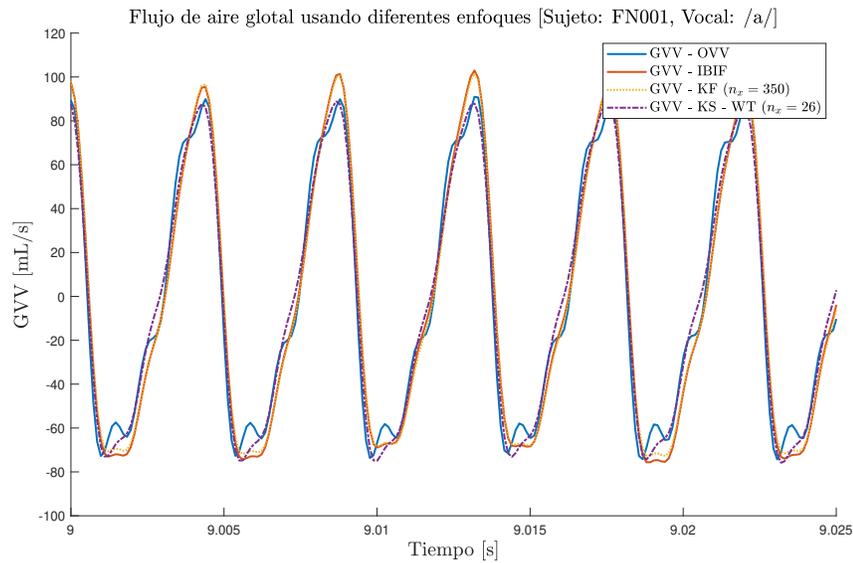
Una solución alternativa a la anterior, es la que se presenta en [65], donde en vez de suponer que la entrada u_t^g es simplemente ruido blanco Gaussiano, ésta se modela como un proceso denominado *Random Walk*. En específico, la ecuación de estado de u_t^g estaría dada por

$$u_{t+1}^g = u_t^g + w_t \quad (4.26)$$

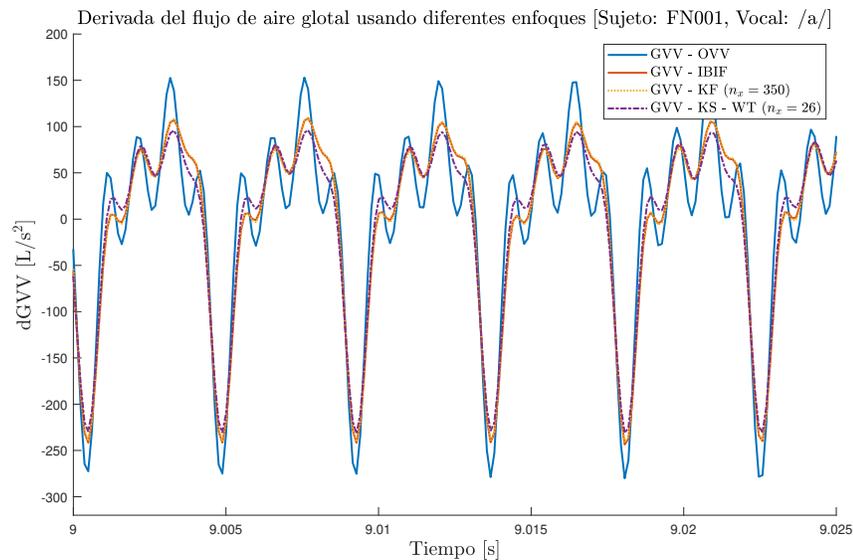
donde $w_t \sim \mathcal{N}(0, \sigma_w^2)$. En este caso, la ecuación anterior modela a u_t^g como un estado que varía sólo gracias a la componente estocástica w_t , por lo que la varianza de ésta debe ser alta (en caso contrario tendríamos malas estimaciones de u_t^g debido a que esta señal no es constante). Con lo anterior, el modelo en variables de estado quedaría igualmente dado por las ecuaciones (4.24)-(4.25) junto a las mismas matrices $\tilde{\mathbf{C}}$ y \mathbf{Q} , con la única diferencia que $\tilde{\mathbf{A}}$ ahora está definida como

$$\tilde{\mathbf{A}}_s = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{R}^{(n_x+1) \times (n_x+1)}. \quad (4.27)$$

En la Fig. 4.15 se muestra el flujo de aire glotal junto a su derivada, obtenido mediante los métodos anteriormente descritos (IBIF subglotal y filtraje de Kalman de alto orden) y mediante el suavizador de Kalman al modelar el flujo de aire glotal como *Random Walk* y considerando una representación en variables de estado de orden 26 con varianzas iguales a $\sigma_w^2 = 10000$ y $\sigma_v^2 = 100$. En la figura podemos ver que las estimaciones son de similar calidad que en la sección 4.4.1. Sin embargo, si consideramos el RMSE cometido al estimar el flujo de aire glotal mediante este esquema, notamos que su valor es de 8.8255 (junto a un NRMSE de 5.4275 %), lo cual es superior al caso anterior. Esto podría cambiar al variar el valor de las varianzas σ_w^2 y σ_v^2 .



(a) Flujo de aire glotal.



(b) Derivada del flujo de aire glotal.

Figura 4.14: Flujo de aire glotal (GVV, en morado) junto a su derivada (dGVV, en morado) obtenidos mediante suavizado de Kalman (KS) y el modelamiento del flujo de aire glotal como ruido blanco (WT) Gaussiano. A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV) y las estimaciones del flujo de aire glotal junto a su derivada obtenidas a partir de IBIF subglotal (en rojo) y a partir del filtraje de Kalman (en amarillo). La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /a/.

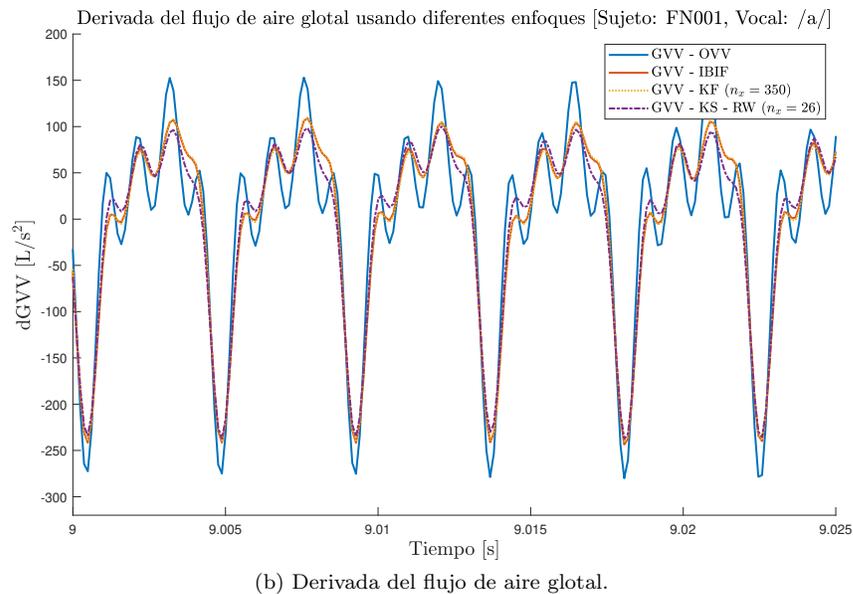
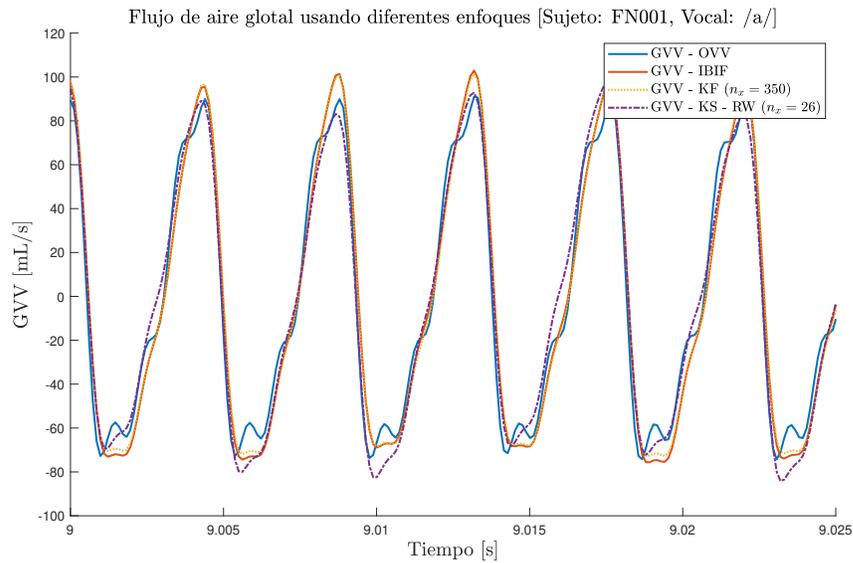


Figura 4.15: Flujo de aire glotal (GVV, en morado) junto a su derivada (dGVV, en morado) obtenidos mediante suavizamiento de Kalman (KS) y el modelamiento del flujo de aire glotal como *Random Walk* (RW). A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV) y las estimaciones del flujo de aire glotal junto a su derivada obtenidas a partir de IBIF subglotal (en rojo) y a partir del filtraje de Kalman (en amarillo). La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /a/.

4.4.3. Modelando la entrada con el modelo de Butterworth (BW)

En las secciones anteriores se ha modelado el flujo de aire glotal de manera simple, suponiendo que no se cuenta con información adicional para estimar la señal en cuestión. Sin embargo, como se vio en la sección 2.2, es posible modelar el flujo de aire glotal u_t^g como la salida de un sistema (representado por un filtro) excitado por un tren de impulsos. Así, podemos utilizar dicho filtro para darle una *forma* más característica a la señal que deseamos estimar. En específico, si utilizamos el modelo de Butterworth presentado en la sección 2.2.1, podemos suponer que u_t^g está dado por

$$u_t^g = \frac{H_0 z^{-2}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} w_t \quad (4.28)$$

donde $w_t \sim \mathcal{N}(0, \sigma_w^2)$ y $\alpha_1 = -(z_1 + z_2)$ y $\alpha_2 = z_1 z_2$ (con respecto a la ec. (2.5)). Para simplificar el numerador, es posible definir el ruido blanco Gaussiano $\tilde{w}_t \sim \mathcal{N}(0, \sigma_w^2)$ de la siguiente manera

$$\tilde{w}_t = H_0 z^{-1} w_t = H_0 w_{t-1}, \quad (4.29)$$

el cual es una versión amplificada y desplazada temporalmente del ruido original w_t . Esto último no supone un problema debido a que en ambos casos u_t^g es la salida de un filtro excitado por ruido blanco Gaussiano cuya varianza es un parámetro a determinar.

Ahora bien, si volvemos a tomar como base el modelo IB subglotal en variables de estado definido por las ec. (4.15)-(4.16) y definimos el vector de estados

$$\tilde{\mathbf{x}}_t^s = \begin{bmatrix} \mathbf{x}_t^s & u_{t-1}^g & u_t^g \end{bmatrix}^\top \quad (4.30)$$

entonces, utilizando la ecuación de diferencias obtenida de la ec. (4.28) y la definición de \tilde{w}_t

$$u_{t+1} + \alpha_1 u_t + \alpha_2 u_{t-1} = \tilde{w}_t \quad (4.31)$$

es posible convertir el sistema determinístico en el siguiente sistema estocástico sin entradas:

$$\tilde{\mathbf{x}}_{t+1}^s = \tilde{\mathbf{A}}_s \tilde{\mathbf{x}}_t^s + \mathbf{w}_t \quad (4.32)$$

$$y_t^a = \tilde{\mathbf{C}}_s \tilde{\mathbf{x}}_t^s + v_t \quad (4.33)$$

donde $\mathbf{w}_t \sim \mathcal{N}_w(\mathbf{0}, \mathbf{Q})$, $v_t \sim \mathcal{N}_v(0, \sigma_v^2)$ y

$$\tilde{\mathbf{A}}_s = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s & \mathbf{0} \\ \mathbf{0} & 0 & 1 \\ \mathbf{0} & -\alpha_2 & -\alpha_1 \end{bmatrix} \in \mathbb{R}^{(n_x+1) \times (n_x+2)} \quad (4.34)$$

$$\tilde{\mathbf{C}}_s = \begin{bmatrix} \mathbf{C}_s & 0 & \mathbf{D}_s \end{bmatrix} \in \mathbb{R}^{1 \times (n_x+2)} \quad (4.35)$$

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_w^2 \end{bmatrix} \in \mathbb{R}^{(n_x+2) \times (n_x+2)}. \quad (4.36)$$

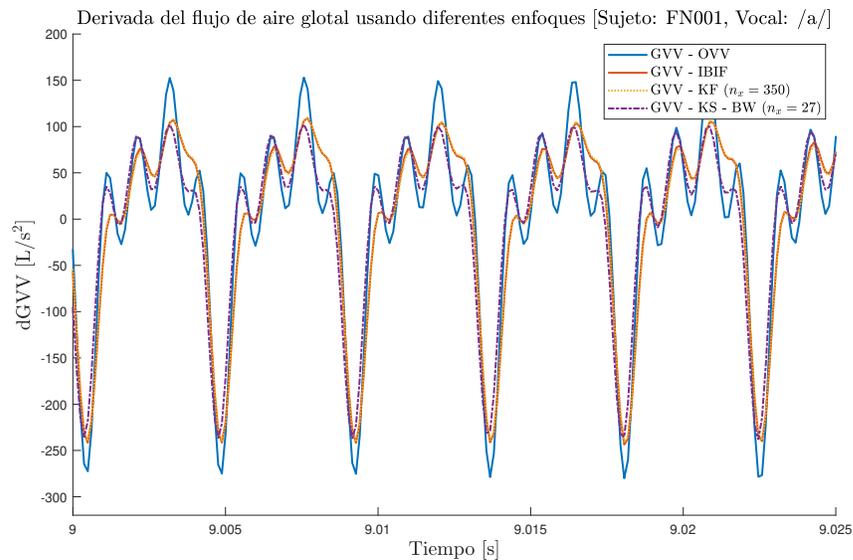
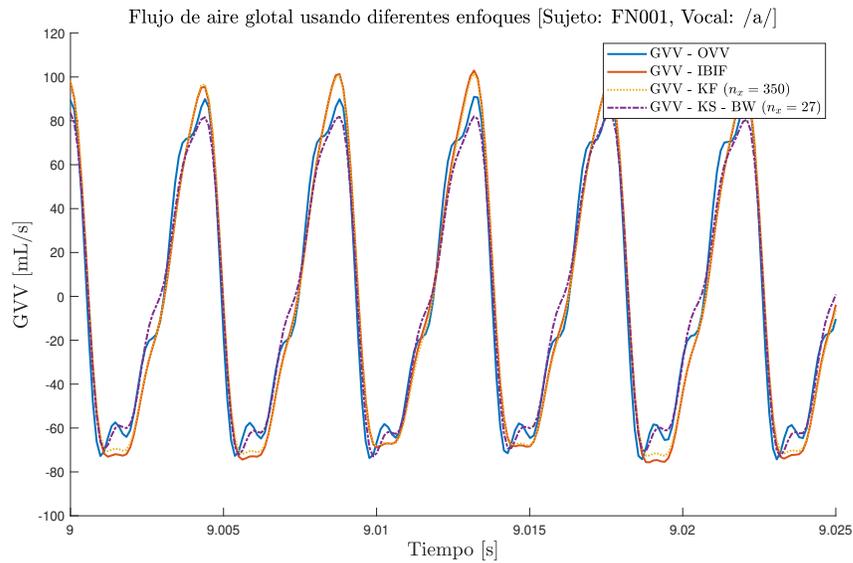


Figura 4.16: Flujo de aire glotal (GVV, en morado) junto a su derivada (dGVV, en morado) obtenidos mediante suavizamiento de Kalman (KS) y el modelamiento del flujo de aire glotal usando el modelo de Butterworth (BW). A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) junto a su derivada (dGVV, en azul) obtenidos mediante el filtraje inverso del flujo de aire oral (OVV) y las estimaciones del flujo de aire glotal junto a su derivada obtenidas a partir de IBIF subglotal (en rojo) y a partir del filtraje de Kalman (en amarillo). La medición del flujo de aire oral y de la aceleración sobre la piel del cuello fueron obtenidas del sujeto de prueba FN001 para la vocal /a/.

En la Fig. 4.16 se muestra el flujo de aire glotal junto a su derivada, obtenido mediante los métodos anteriormente descritos (IBIF subglotal y filtraje de Kalman de alto orden) y mediante el suavizador de Kalman al modelar el flujo de aire glotal usando el modelo de Butterworth y considerando una representación en variables de estado de orden 27 con varianzas iguales a $\sigma_w^2 = 1000$ y $\sigma_v^2 = 100$. En la figura podemos ver que las estimaciones son de similar calidad que en las secciones 4.4.1 y 4.4.2. Lo anterior puede ser corroborado al obtener el RMSE cometido al estimar el flujo de aire glotal mediante este modelo, el cual ahora tiene un valor de 8.1337 (y un NRMSE de 4.9546 %), el cual es menor que en el caso del modelo RW pero mayor que en el caso del modelo WT.

4.5. Medida de la incertidumbre

Una de las ventajas de estimar el flujo de aire glotal mediante un enfoque Bayesiano, como lo es el de filtraje de Kalman o suavizamiento de Kalman, es que resuelve el problema de filtrado inverso de una forma probabilística. Esto último significa que en realidad no se estima un valor puntual de la señal a estimar, sino que se estima su función de probabilidad. En específico, recordemos que tanto el filtro de Kalman como el suavizador, buscan estimar los estados de un sistema, los cuales están expuestos a ruido, a partir de las mediciones de la salida, que también poseen ruido. Lo anterior se realiza de manera óptima, en el sentido de minimizar la varianza del error de estimación o el error cuadrático medio (MSE, por sus siglas en inglés), esto es, $\mathbf{P}_t = \mathbb{E}\{(\mathbf{x}_t - \hat{\mathbf{x}}_t)(\mathbf{x}_t - \hat{\mathbf{x}}_t)^\top\}$, donde $\hat{\mathbf{x}}_t$ es la estimación de \mathbf{x}_t .

Un hecho importante es que, bajo la suposición de que el sistema es lineal y los ruidos son Gaussianos, entonces tanto el filtro como el suavizador de Kalman son los mejores estimadores en el sentido de minimizar la varianza del error de estimación o MSE. Con lo anterior, la función de probabilidad de \mathbf{x}_t también es Gaussiana y por tanto basta con conocer la evolución de la media y de la varianza para describirla completamente.

En las secciones anteriores se ha utilizado el *valor más probable* o, en concreto, la media $\hat{\mathbf{x}}_t = \mathbb{E}\{\mathbf{x}_t\}$ como estimación del flujo de aire glotal. Sin embargo, como ahora estamos viendo el problema desde una perspectiva probabilística, existe una *función de distribución de probabilidad* de \mathbf{x}_t en torno al valor $\hat{\mathbf{x}}_t$. Esta probabilidad o *medida de incertidumbre* es valiosa en un ambiente ambulatorio, donde no se tiene una señal de referencia para saber qué tan buenas están siendo las estimaciones y además es una característica que no es posible extraer con el enfoque IBIF.

En efecto, una métrica para cuantificar esta incertidumbre es precisamente la varianza del error de estimación contenida en la matriz \mathbf{P}_t . Así, con el enfoque de suavizamiento de Kalman es posible utilizar como estimación puntual la esperanza $\hat{\mathbf{x}}_t = \mathbb{E}\{\mathbf{x}_t\}$, al mismo tiempo de tener una medida del error que se estaría cometiendo en las estimaciones.

Para ser más concretos, consideremos la Fig. 4.17. En esta figura se muestra (en azul) la señal de referencia obtenida mediante el flujo de aire oral. Al mismo tiempo, la línea de puntos y trazos morada, representa el valor más probable del flujo de aire glotal según el enfoque de suavizamiento de Kalman. Junto a esta línea, se dibuja una banda morada, que representa, para cada t , un intervalo donde existe la posibilidad de que la señal a estimar se encuentre realmente. En específico, la distancia entre la media y la banda superior es de 2 desviaciones estándar; la misma distancia existe entre la media y la banda inferior. Con lo anterior, existe aproximadamente un 95 % de probabilidades de que la señal a estimar se encuentre dentro de la banda.

Aquí es importante mencionar que estas estimaciones y probabilidades son óptimas cuando se utiliza el modelo verdadero del sistema para hacer el suavizamiento de Kalman.

Dado que se utilizan datos reales, no se conoce el modelo verdadero que los origina. Por tanto, no hay garantía de que la estimación sea óptima y en consecuencia la señal de referencia no siempre se encontrará dentro de la banda morada.

A pesar de lo anterior, en un ambiente ambulatorio el conocimiento de la varianza del error de estimación permite generar un intervalo de confianza, además de cuantificar la incertidumbre que existe en las estimaciones.

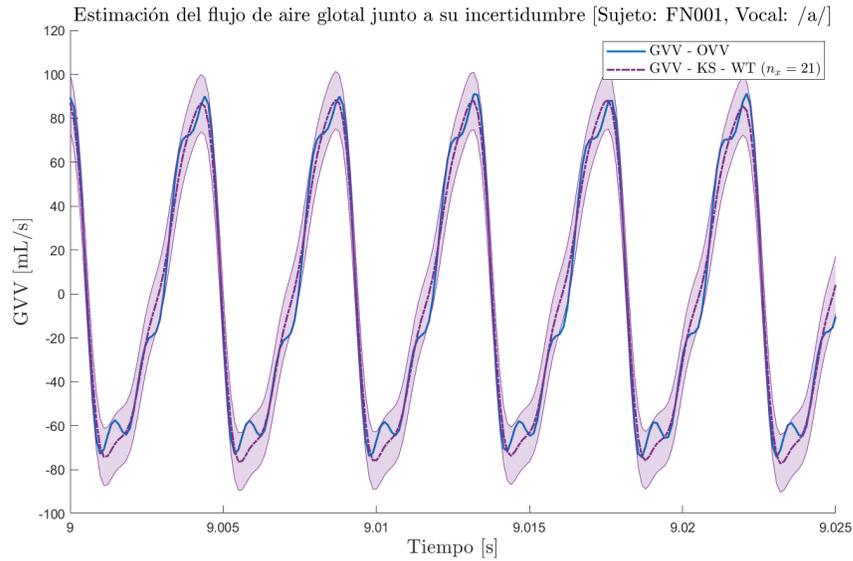


Figura 4.17: Estimación (valor esperado) del flujo de aire glotal (GVV, línea de puntos y trazos morada) obtenido mediante suavizamiento de Kalman (KS) usando el modelo de ruido blanco (WT), junto a una banda que refleja la varianza del error de estimación. Esta banda tiene un ancho total de 4 desviaciones estándar. A modo de comparación, se incluye como señal de referencia el flujo de aire glotal (GVV, en azul) obtenido a partir del flujo de aire oral.

RESULTADOS NUMÉRICOS

En el capítulo 3 se mostró cómo obtener una representación en variables de estado del modelo basado en impedancia subglotal, mientras que en el capítulo 4 se mostró cómo utilizar esta representación para estimar el flujo de aire glotal mediante un suavizador de Kalman utilizando tres modelos distintos para la entrada u_t^g . El presente capítulo tiene como objetivo exhibir resultados numéricos del enfoque aquí desarrollado, además de realizar una comparación con los enfoques anteriores. En específico, se comienza explicando cómo seleccionar los parámetros que definen al suavizador de Kalman. Luego, se realiza una comparación entre las estimaciones del flujo de aire glotal al calcular la raíz del error cuadrático medio utilizando como referencia el flujo de aire glotal obtenido a partir del flujo de aire oral. Posteriormente, se comparan las características aerodinámicas del flujo de aire glotal que son fundamentales para la evaluación de la hiperfunción vocal. Finalmente, se realiza una comparación relacionada al tiempo de cómputo requerido por cada enfoque.

5.1. Elección de parámetros

Estimar el flujo de aire glotal mediante el enfoque de suavizamiento de Kalman, requiere de la elección de ciertos *parámetros de diseño*, los cuales definen al suavizador. En específico, como se vio en la sección 4.4, los tres modelos utilizados para representar a la entrada u_t^g , los cuales son: el modelo de ruido blanco (WT), el modelo *random walk* (RW) y el modelo de Butterworth (BW), requieren definir la varianza del ruido de proceso σ_w^2 y la varianza del ruido de salida σ_v^2 . Por lo anterior, y debido a que la elección de estos parámetros define la calidad de las estimaciones, es necesario definir un procedimiento para encontrarlos.

Antes de comenzar, es necesario recordar del capítulo 3 que los modelos en variables de estado que mejor representaban al modelo IB subglotal, fueron aquellos obtenidos mediante la identificación en el dominio de la frecuencia. En específico, de los órdenes analizados, los que mejor dieron resultados, al mismo tiempo de ser poco complejos, fueron los de orden 15, 20, 25, 30 y 35. Por lo anterior, aquí analizaremos esos 5 modelos en variables de estado. Como además tenemos tres formas de modelar el ruido u_t^g , lo anterior nos da como resultado 15 posibles candidatos de modelos en variables de estado sobre los cuales aplicar el suavizador de Kalman.

Ahora bien, una métrica para cuantificar el error cometido por las estimaciones es la raíz del error cuadrático medio (RMSE) obtenido al comparar las estimaciones con la señal de referencia (esto es, el flujo de aire glotal obtenido a partir del flujo de aire oral). Lo anterior tiene sentido considerando que la idea es obtener estimaciones lo más cercanas posibles a la señal de referencia cuando son comparadas temporalmente. En efecto, en [23] se utiliza dicha métrica para encontrar el par (σ_w^2, σ_v^2) . El procedimiento consiste en obtener estimaciones del flujo de aire glotal utilizando distintos valores (σ_w^2, σ_v^2) y seleccionar el par que menor RMSE entregue.

En el presente trabajo se escoge un procedimiento ligeramente distinto. Como el objetivo principal es obtener buenas estimaciones de las características aerodinámicas del flujo de aire glotal, se plantea utilizar una métrica que entregue información sobre dichas características. Así, aquí se utilizará una métrica denominada *error absoluto medio ponderado* (WMAE, por sus siglas en inglés), la cual está basada principalmente en la métrica NWAEE, introducida en [32]. La métrica WMAE está definida como

$$\text{WMAE}(\sigma_w^2, \sigma_v^2) = \frac{1}{2} \sum_{i=1}^2 \left(\frac{1}{N} \sum_{t=1}^N |\Delta^{(i-1)} \tilde{u}_t^g - \Delta^{(i-1)} \hat{u}_t^g| \right) \quad (5.1)$$

donde \tilde{u}_t^g es la señal de referencia, \hat{u}_t^g es el flujo de aire glotal estimado (y sincronizado temporalmente con la señal de referencia) y $\Delta^{(i-1)}$ es el operador derivada de orden $(i-1)$. Para sincronizar o alinear temporalmente el flujo estimado \hat{u}_t^g con la señal de referencia \tilde{u}_t^g , se utiliza el enfoque utilizado en [23, 32], donde se calcula la correlación cruzada entre ambas señales, para luego desplazar una señal con respecto a la otra utilizando el corrimiento temporal que entrega la mayor correlación. Al utilizar la métrica WMAE, la idea es cuantificar el error absoluto tanto en el flujo de aire glotal como en su derivada, debido a que ambas señales son utilizadas para calcular las características aerodinámicas.

Un ejemplo de los datos utilizados en el presente trabajo se muestra en la Fig. 5.1, donde se exhibe el flujo de aire glotal obtenido a partir del flujo de aire oral para el sujeto FN001 y la vocal /a/. En este caso, el flujo de aire oral fue muestreado a una frecuencia de 8192 Hz, por un total de 13.1083 segundos equivalentes a 107 383 muestras. Aquí, el sujeto pronunció la vocal /a/ 5 veces. Así, para obtener el WMAE, se comienza encontrando las ventanas de voz, para luego comparar el flujo estimado con la señal de referencia en cada ventana de voz y finalmente obtener la media entre estos valores.

En la Fig. 5.2 se muestra un mapa de colores donde se muestra el WMAE para distintas combinaciones de los parámetros, los cuales se hicieron variar en múltiplos de 10 en los siguientes rangos: $\sigma_w^2 \in [10^{-4}, 10^6]$ y $\sigma_v^2 \in [10^{-6}, 10^4]$. En este caso, podemos ver que el mínimo se alcanza en todo el área azul oscura. Además, es posible notar que al subir por las diagonales los valores se mantienen constantes, lo que significa que lo que importa en realidad es la relación entre σ_w^2 y σ_v^2 más que sus valores particulares. Para eliminar la igualdad producida, se propone utilizar como segunda métrica el RMSE:

$$\text{RMSE}(\sigma_w^2, \sigma_v^2) = \sqrt{\frac{1}{N} \sum_{t=1}^N (\tilde{u}_t^g - \hat{u}_t^g)^2}. \quad (5.2)$$

Una vez realizado el procedimiento para obtener el par (σ_w^2, σ_v^2) , en la Tabla 5.1 se muestran las combinaciones de dichos parámetros elegidos para cada sujeto, orden de sistema y modelo de u_t^g utilizado. En la tabla, existen valores que no están dados, lo cual es debido a que para dicha combinación sujeto-orden-modelo, no existía un par (σ_w^2, σ_v^2) que hiciera

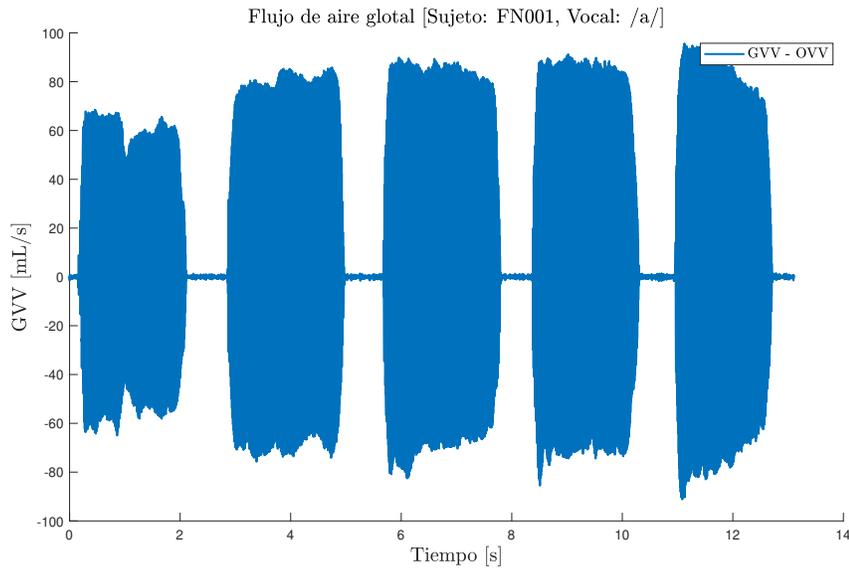


Figura 5.1: Ejemplo de datos utilizados. Flujo de aire glotal obtenido a partir del flujo de aire oral para el sujeto FN001 utilizando la vocal /a/.

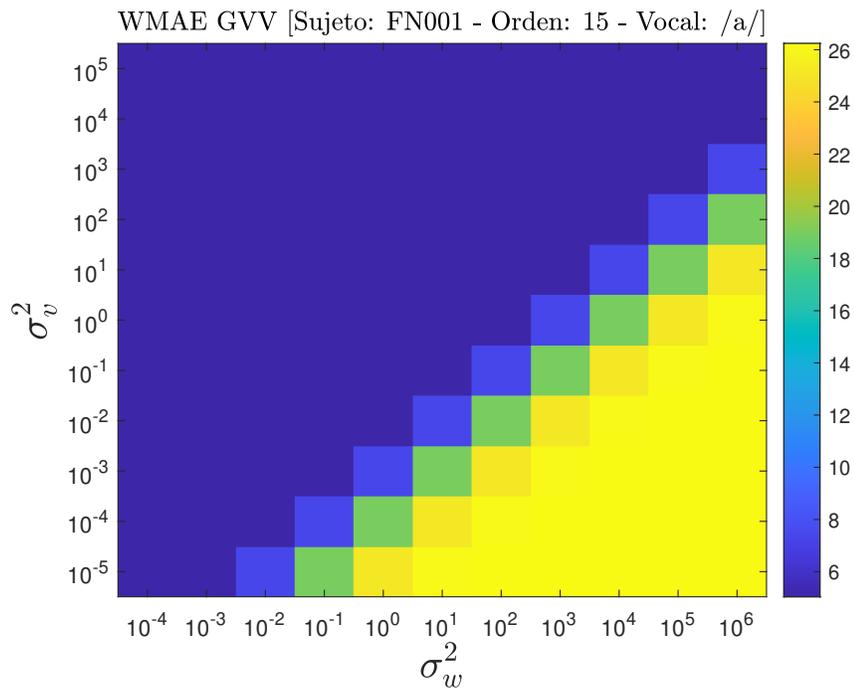


Figura 5.2: Valores de WMAE para distintas combinaciones (σ_w^2, σ_v^2) en el caso del sujeto FN001 al pronunciar la vocal /a/ y considerando el modelo WT de orden 15.

que el suavizador de Kalman convergiera. Así, y debido a que los órdenes 30 y 35 fueron los que mayor problemas de convergencia tenían, de aquí en adelante no se considerarán para efectos de comparación.

Método	n_x	Parámetros (σ_w^2, σ_v^2)					
		Sujeto					
		FN001	FN003	FN006	FP003	FP004	FP008
KS (WT)	15(+1)	(10 ⁴ , 10 ²)	(10 ⁻⁴ , 10 ⁻⁶)	(10 ⁶ , 10 ⁴)	(10 ⁵ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	20(+1)	(10 ³ , 10 ²)	(10 ⁻¹ , 10 ⁻³)	(10 ³ , 10 ²)	(10 ⁵ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	25(+1)	(10 ³ , 10 ²)	(10 ⁻⁴ , 10 ⁻⁶)	(10 ¹ , 10 ¹)	(10 ⁴ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	30(+1)	–	–	(10 ¹ , 10 ¹)	–	–	–
	35(+1)	(10 ⁵ , 10 ³)	–	–	–	(10 ⁶ , 10 ⁴)	(10 ⁶ , 10 ⁴)
KS (RW)	15(+1)	(10 ⁵ , 10 ²)	(10 ⁵ , 10 ²)	(10 ⁴ , 10 ²)	(10 ⁵ , 10 ²)	(10 ⁻¹ , 10 ⁻¹)	(10 ⁻¹ , 10 ⁻¹)
	20(+1)	(10 ⁵ , 10 ²)	(10 ⁵ , 10 ²)	(10 ⁴ , 10 ²)	(10 ⁵ , 10 ²)	(10 ¹ , 10 ²)	(10 ¹ , 10 ²)
	25(+1)	(10 ⁴ , 10 ²)	(10 ⁵ , 10 ²)	–	(10 ⁵ , 10 ²)	(10 ¹ , 10 ²)	(10 ¹ , 10 ²)
	30(+1)	–	–	(10 ⁻³ , 10 ⁻⁴)	–	–	–
	35(+1)	(10 ⁴ , 10 ²)	–	–	–	(10 ⁵ , 10 ⁴)	(10 ⁵ , 10 ⁴)
KS (BW)	15(+2)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ⁻³ , 10 ⁻⁵)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	20(+2)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ⁶ , 10 ⁴)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	25(+2)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ¹ , 10 ¹)	(10 ³ , 10 ²)	(10 ³ , 10 ²)	(10 ³ , 10 ²)
	30(+2)	–	–	(10 ³ , 10 ²)	–	–	–
	35(+2)	(10 ⁵ , 10 ³)	–	–	–	(10 ⁵ , 10 ⁴)	(10 ⁵ , 10 ⁴)

Tabla 5.1: Parámetros (σ_w^2, σ_v^2) del suavizador de Kalman elegidos en base al menor error absoluto medio ponderado obtenido para la vocal /a/. Los guiones representan casos donde el suavizador de Kalman diverge.

5.2. Flujo de aire glotal

Una vez elegidos los parámetros σ_w^2 y σ_v^2 es posible comparar las estimaciones del flujo de aire glotal obtenidas mediante los diferentes enfoques. La primera métrica que se utiliza como comparación es el RMSE. El procedimiento para calcular el RMSE es el mismo que para calcular el WMAE, donde primero se encuentran las ventanas de voz y luego se compara el flujo estimado y sincronizado con la señal de referencia.

Una segunda métrica útil para comparar los enfoques, la cual está relacionada a la primera, es la *raíz del error cuadrático medio normalizada* (NRMSE, por sus siglas en inglés), la cual aquí definimos como

$$\text{NRMSE} = \frac{\text{RMSE}}{\tilde{u}_{\text{máx}}^g - \tilde{u}_{\text{mín}}^g} = \frac{1}{\tilde{u}_{\text{máx}}^g - \tilde{u}_{\text{mín}}^g} \sqrt{\frac{1}{N} \sum_{t=1}^N (\tilde{u}_t^g - \hat{u}_t^g)^2}, \quad (5.3)$$

y cuyo objetivo es cuantificar el error con respecto al valor *peak-to-peak* de la señal.

A continuación, en las Tablas 5.2 y 5.3 se muestran, respectivamente, el RMSE y el NRMSE cometido al estimar el flujo de aire glotal a partir de la vocal /a/ para cada sujeto y utilizando los diferentes enfoques. Para las estimaciones obtenidas mediante el enfoque de filtraje de Kalman de alto orden se ha utilizado $\sigma_w^2 = 100$ y $\sigma_v^2 = 1$, tal como se propuso en [23], mientras que para el enfoque de suavizamiento de Kalman se utilizan los valores de la Tabla 5.1. En las Tablas 5.2 y 5.3, WT hace referencia al modelo de ruido blanco, RW al modelo *random walk* y BW al modelo de Butterworth. Los valores que no están dados son debido a que en tal caso el suavizador de Kalman divergió.

En la Tabla 5.2 podemos notar que el enfoque de suavizamiento de Kalman que mejor resultados entrega es el que utiliza el modelo de Butterworth, donde en todos los casos se entregan valores RMSE menores que en el caso de IBIF y de KF. Posteriormente viene el modelo de ruido blanco, donde dependiendo del orden del sistema, el RMSE es ligeramente menor o mayor que en el caso de IBIF y de KF. Finalmente, el modelo *random walk* es el que presenta mayor RMSE con respecto a los demás enfoques para todos los órdenes. Por otro lado, en la Tabla 5.3 podemos ver que al comparar el NRMSE promedio (última columna), el enfoque de KS presenta mejores resultados que IBIF y KF en la mayoría de casos. Lo anterior indica que el enfoque de suavizamiento de Kalman entrega, en promedio, menores errores porcentuales con respecto a IBIF y KF. El hecho de que existan casos donde el NRMSE promedio invierta la tendencia que se tenía con el RMSE promedio es explicado al considerar que cada sujeto posee un valor *peak-to-peak* diferente del flujo de aire glotal.

Ahora bien, sería valioso analizar el caso donde las señales medidas son obtenidas a partir de una vocal distinta a la utilizada para calibrar el modelo IB subglotal (la cual, en el presente trabajo, es la vocal /a/). Por lo anterior, en las Tablas 5.4 y 5.5 se muestran, respectivamente, el RMSE y el NRMSE cometido considerando ahora la vocal /i/. En este caso, podemos notar un aumento general en cuanto al RMSE y al NRMSE con respecto a las tablas anteriores, lo que resulta natural considerando que los parámetros del modelo IB subglotal fueron obtenidos mediante un esquema de calibración que utilizaba como referencia el flujo de aire glotal para la vocal /a/. A pesar de lo anterior, el esquema de suavizamiento de Kalman presenta, para todos sus casos, un menor error con respecto a los esquemas IBIF y KF. Al mismo tiempo, el enfoque KF también presenta un menor error con respecto al esquema IBIF. Lo anterior confirma la hipótesis de que el enfoque Bayesiano propuesto en esta tesis permite obtener estimaciones similares o mejores que las documentadas previamente en la literatura.

5.3. Características aerodinámicas

Las características aerodinámicas extraídas del flujo de aire glotal son de vital importancia para la evaluación de la función vocal. Por lo anterior, en esta sección se comparan los diferentes enfoques para estimar las características aerodinámicas del flujo de aire glotal.

El procedimiento para obtener las características aerodinámicas es el mismo que se ha utilizado en estudios anteriores, esto es, primero se encuentran las ventanas con voz y luego se dividen en ventanas de 50 ms. Posteriormente, las medidas aerodinámicas son obtenidas en cada ventana de 50 ms, y finalmente estos valores son promediados para obtener una estimación de la medida en cuestión.

Para comparar los enfoques, analizaremos el error relativo cometido al estimar cada característica, el cual es definido como:

RMSE - Vocal: /a/								
Método	n_x	Sujeto						$\overline{\text{RMSE}}$
		FN001	FN003	FN006	FP003	FP004	FP008	
IBIF		8.3881	9.2973	7.4145	15.3621	24.9531	4.5120	11.6545
KF	350	8.1494	8.8906	7.4065	16.2904	24.3745	4.5429	11.6091
KS (WT)	15(+1)	8.4132	8.4986	7.3988	15.5854	25.9610	4.3856	11.7071
	20(+1)	8.3390	8.6041	7.2384	15.4886	25.9607	4.3744	11.6675
	25(+1)	8.0694	8.2999	8.0481	15.2679	24.9409	4.3951	11.5036
KS (RW)	15(+1)	8.5089	8.5519	7.4559	15.5870	26.2555	4.4002	11.7932
	20(+1)	8.6850	8.6641	7.4948	15.4908	25.7749	4.3995	11.7515
	25(+1)	8.8255	8.5619	–	15.3502	24.8343	4.4878	12.4119
KS (BW)	15(+2)	8.5346	6.4252	7.2960	14.2913	25.0591	4.5118	11.0197
	20(+2)	8.3072	6.4790	7.2239	14.0572	24.9436	4.4870	10.9163
	25(+2)	8.1337	6.3005	7.3999	13.9032	24.1465	4.5449	10.7381

Tabla 5.2: Raíz del error cuadrático medio obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /a/. La última columna presenta el promedio de cada fila.

NRMSE - Vocal: /a/								
Método	n_x	Sujeto						$\overline{\text{NRMSE}}$
		FN001	FN003	FN006	FP003	FP004	FP008	
IBIF		5.1020 %	6.5589 %	5.8734 %	4.5152 %	7.3443 %	5.3918 %	5.7976 %
KF	350	4.9481 %	6.2550 %	5.9069 %	4.8017 %	7.1683 %	5.4415 %	5.7536 %
KS (WT)	15(+1)	5.1734 %	5.9843 %	5.8517 %	4.5815 %	7.5968 %	5.1458 %	5.7222 %
	20(+1)	5.1237 %	6.0603 %	5.7387 %	4.5520 %	7.6014 %	5.1390 %	5.7025 %
	25(+1)	4.9486 %	5.8375 %	6.4207 %	4.4887 %	7.2973 %	5.1753 %	5.6947 %
KS (RW)	15(+1)	5.2337 %	6.0252 %	5.8972 %	4.5820 %	7.6895 %	5.1639 %	5.7652 %
	20(+1)	5.3480 %	6.1061 %	5.9392 %	4.5526 %	7.5385 %	5.1699 %	5.7757 %
	25(+1)	5.4275 %	6.0362 %	–	4.5127 %	7.2585 %	5.2904 %	5.7051 %
KS (BW)	15(+2)	5.1999 %	4.5135 %	5.7805 %	4.2060 %	7.3209 %	5.3393 %	5.3933 %
	20(+2)	5.0693 %	4.5511 %	5.7325 %	4.1359 %	7.2895 %	5.3177 %	5.3493 %
	25(+2)	4.9546 %	4.4217 %	5.8956 %	4.0932 %	7.0523 %	5.3997 %	5.3028 %

Tabla 5.3: Raíz del error cuadrático medio normalizado obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /a/. La última columna presenta el promedio de cada fila.

RMSE - Vocal: /i/								
Método	n_x	Sujeto						$\overline{\text{RMSE}}$
		FN001	FN003	FN006	FP003	FP004	FP008	
IBIF		10.9623	11.9543	13.5481	32.0535	33.7576	8.2014	18.4129
KF	350	10.7145	11.4343	12.8564	30.5154	33.1365	7.9136	17.7618
KS (WT)	15(+1)	11.4257	9.9772	13.6244	30.8731	30.5976	6.7308	17.2048
	20(+1)	11.1775	10.0810	13.0902	30.8939	30.7109	6.7714	17.1208
	25(+1)	10.9286	9.7658	11.6709	30.2824	29.6072	6.6998	16.4925
KS (RW)	15(+1)	11.5104	10.1373	13.6853	30.8826	31.3075	6.8379	17.3935
	20(+1)	11.4934	10.2470	13.4945	30.9044	28.3343	6.8840	16.8930
	25(+1)	11.5751	10.0860	–	30.4644	27.3691	6.8550	17.2699
KS (BW)	15(+2)	11.9810	10.7465	13.3336	29.6893	27.7373	6.5421	16.6716
	20(+2)	11.7050	10.8113	13.0859	29.7048	27.7723	6.5864	16.6110
	25(+2)	11.5913	10.6577	12.3249	28.9470	26.8327	6.5735	16.1545

Tabla 5.4: Raíz del error cuadrático medio obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /i/. La última columna presenta el promedio de cada fila.

NRMSE - Vocal: /i/								
Método	n_x	Sujeto						$\overline{\text{NRMSE}}$
		FN001	FN003	FN006	FP003	FP004	FP008	
IBIF		6.3121 %	12.5573 %	12.0773 %	20.1106 %	11.1721 %	12.3768 %	12.4344 %
KF	350	6.1668 %	12.0082 %	11.6040 %	18.9760 %	10.9372 %	11.9445 %	11.9395 %
KS (WT)	15(+1)	6.5839 %	10.4744 %	12.0925 %	18.9528 %	10.1800 %	10.1414 %	11.4042 %
	20(+1)	6.4393 %	10.5843 %	11.7158 %	18.9545 %	10.2256 %	10.2068 %	11.3544 %
	25(+1)	6.2944 %	10.2510 %	10.7469 %	18.6494 %	9.9118 %	10.1070 %	10.9934 %
KS (RW)	15(+1)	6.6339 %	10.6431 %	12.1654 %	18.9632 %	10.4089 %	10.3004 %	11.5192 %
	20(+1)	6.6287 %	10.7593 %	12.1860 %	18.9675 %	9.5959 %	10.3741 %	11.4186 %
	25(+1)	6.6772 %	10.5892 %	–	18.9934 %	9.3357 %	10.3395 %	11.1870 %
KS (BW)	15(+2)	6.8882 %	11.2809 %	11.8780 %	18.2444 %	9.4146 %	9.8683 %	11.2624 %
	20(+2)	6.7324 %	11.3502 %	11.7279 %	18.2517 %	9.4390 %	9.9379 %	11.2399 %
	25(+2)	6.6645 %	11.1892 %	11.1784 %	17.7040 %	9.1810 %	9.9265 %	10.9739 %

Tabla 5.5: Raíz del error cuadrático medio normalizado obtenido a partir de la comparación en el tiempo de la señal de referencia con el flujo de aire glotal estimado a partir de la vocal /i/. La última columna presenta el promedio de cada fila.

$$e_r = \left| \frac{\tilde{y} - \hat{y}}{\tilde{y}} \right| \quad (\%) \quad (5.4)$$

donde \tilde{y} es el valor de la característica aerodinámica obtenida a partir de la señal de referencia y \hat{y} es el valor obtenido a partir del flujo de aire estimado mediante alguno de los diferentes esquemas.

Las Tablas 5.6 y 5.7 muestran la media junto a la desviación estándar de los errores relativos cometidos al estimar las diferentes características aerodinámicas para los distintos sujetos considerando la vocal /a/. En el caso del enfoque KS-RW de orden 25, se han omitido los resultados debido a que dicho enfoque diverge al intentar obtener estimaciones para el sujeto FN006. De ambas tablas, podemos ver que en general el esquema IBIF entrega mejores resultados de ACFL y MFDR, en comparación con los enfoques KS, mientras que lo contrario ocurre para el caso de H1-H2, HRF, SQ y NAQ. En específico, podemos ver que las características obtenidas en el dominio de la frecuencia, tales como H1-H2 y HRF, son mejores para el enfoque KS que utiliza el modelo de Butterworth, lo cual tiene sentido considerando que dicho modelo intenta describir al flujo de aire glotal utilizando sus propiedades espectrales.

Para simplificar las comparaciones, se seleccionan los dos mejores casos del esquema KS, considerando como mejor a aquel que tenga la mayor cantidad de características aerodinámicas mejor estimadas con respecto al enfoque IBIF. En específico, en el caso del esquema KS que usa el modelo de ruido blanco se selecciona el que utiliza el modelo de orden 20, al igual que en el caso donde se usa el modelo de Butterworth. Ambos casos seleccionados, son destacados en color en las tablas. Si ahora realizamos una comparación con las características obtenidas mediante IBIF, tenemos que el enfoque KS - WT entrega menores errores en 4 casos: H1-H2, F0, SQ y NAQ. Por otro lado, el enfoque KS - BT entrega menores errores en 5 casos: H1-H2, HRF, F0, SQ y NAQ. Por su parte, el esquema IBIF posee un menor error relativo promedio en el caso de ACFL, MFDR y OQ, en comparación con los dos mejores casos del enfoque de suavizamiento de Kalman.

Para comparar los errores cuando la vocal pronunciada es diferente a la utilizada para construir el modelo basado en impedancia subglotal, en las Tablas 5.8 y 5.9 se muestran la media junto a la desviación estándar de los errores relativos considerando la vocal /i/. En este nuevo caso, podemos ver que existe un aumento general en los errores relativos en comparación con el caso de la vocal /a/, lo que es esperable considerando la forma en que fue calibrado el modelo basado en impedancia subglotal. Sin embargo, si consideramos los dos mejores casos seleccionados anteriormente, podemos ver que, al menos en el enfoque KS que utiliza el modelo de ruido blanco, existe una notable mejora en comparación con el esquema IBIF subglotal. En efecto, 6 características son obtenidas con un error relativo promedio inferior que en el caso de IBIF, los cuales son: ACFL, MFDR, H1-H2, HRF, F0 y NAQ. En el caso del enfoque KS que utiliza el modelo de Butterworth, sólo son 3 los casos en que se obtienen menores errores con respecto a IBIF: ACFL, MFDR y F0. En resumen, es posible concluir que el enfoque KS que utiliza el modelo de ruido blanco, entrega estimaciones comparativamente mejores que el esquema IBIF si se consideran las vocales /a/ e /i/, mientras que el que utiliza el modelo de Butterworth entrega un rendimiento similar con respecto al esquema IBIF.

El análisis anterior estaba basado en comparar el valor promedio de cada característica entre todas las ventanas de 50 ms. Esto quiere decir que si se tenían, digamos, 200 ventanas, con 200 distintos valores, sólo importaba el valor promedio. Un análisis interesante, sería

Error relativo en características aerodinámicas - Vocal: /a/					
Método	n_x	Características Aerodinámicas			
		ACFL	MFDR	H1-H2	HRF
IBIF		3.150 % ± 2.311	12.545 % ± 5.355	18.114 % ± 20.641	29.675 % ± 20.641
KF	350	4.556 % ± 2.686	14.082 % ± 5.225	20.371 % ± 20.148	28.344 % ± 20.148
KS (WT)	15(+1)	5.868 % ± 4.663	15.434 % ± 5.609	18.286 % ± 20.477	30.066 % ± 20.477
	20(+1)	5.900 % ± 3.924	15.733 % ± 5.322	17.885 % ± 21.705	30.839 % ± 21.705
	25(+1)	7.890 % ± 3.813	17.476 % ± 3.715	18.421 % ± 21.100	30.370 % ± 21.100
KS (RW)	15(+1)	5.407 % ± 4.631	15.224 % ± 5.529	18.061 % ± 20.766	29.763 % ± 20.766
	20(+1)	5.729 % ± 4.344	16.227 % ± 5.388	17.545 % ± 22.048	31.396 % ± 22.048
	25(+1)	–	–	–	–
KS (BW)	15(+2)	8.446 % ± 4.103	13.593 % ± 4.992	17.007 % ± 20.524	20.024 % ± 20.524
	20(+2)	8.025 % ± 3.855	13.460 % ± 4.741	16.876 % ± 21.348	20.703 % ± 21.348
	25(+2)	9.066 % ± 2.844	14.154 % ± 3.991	17.113 % ± 20.805	20.062 % ± 20.805

Tabla 5.6: Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /a/ (1 de 2).

Error relativo en características aerodinámicas - Vocal: /a/					
Método	n_x	Características Aerodinámicas			
		F0	OQ	SQ	NAQ
IBIF		0.132 % ± 0.202	8.220 % ± 5.159	15.192 % ± 7.220	14.007 % ± 7.220
KF	350	0.171 % ± 0.140	8.425 % ± 5.654	16.982 % ± 6.746	13.167 % ± 6.746
KS (WT)	15(+1)	0.130 % ± 0.236	11.533 % ± 8.457	11.730 % ± 13.209	10.640 % ± 13.209
	20(+1)	0.130 % ± 0.237	11.792 % ± 7.933	11.450 % ± 12.946	10.996 % ± 12.946
	25(+1)	0.157 % ± 0.230	11.345 % ± 7.241	11.748 % ± 12.950	10.789 % ± 12.950
KS (RW)	15(+1)	0.155 % ± 0.233	11.525 % ± 8.501	11.579 % ± 12.846	10.891 % ± 12.846
	20(+1)	0.158 % ± 0.232	11.738 % ± 8.094	11.012 % ± 12.073	11.977 % ± 12.073
	25(+1)	–	–	–	–
KS (BW)	15(+2)	0.108 % ± 0.130	10.909 % ± 8.005	13.135 % ± 12.065	5.212 % ± 12.065
	20(+2)	0.109 % ± 0.131	11.052 % ± 7.465	12.456 % ± 11.725	5.542 % ± 11.725
	25(+2)	0.109 % ± 0.127	10.418 % ± 6.804	11.788 % ± 11.979	5.175 % ± 11.979

Tabla 5.7: Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /a/ (2 de 2).

Error relativo en características aerodinámicas - Vocal: /i/					
Método	n_x	Características Aerodinámicas			
		ACFL	MFDR	H1-H2	HRF
IBIF		39.650 % ± 26.809	45.659 % ± 28.879	25.110 % ± 14.161	24.306 % ± 14.161
KF	350	39.087 % ± 27.005	45.717 % ± 29.160	27.206 % ± 14.946	27.453 % ± 14.946
KS (WT)	15(+1)	34.219 % ± 25.053	39.610 % ± 27.278	25.699 % ± 12.855	24.153 % ± 12.855
	20(+1)	33.843 % ± 25.630	39.336 % ± 27.823	24.417 % ± 13.267	23.536 % ± 13.267
	25(+1)	31.302 % ± 26.914	37.128 % ± 29.123	27.316 % ± 12.852	27.419 % ± 12.852
KS (RW)	15(+1)	34.710 % ± 24.967	39.890 % ± 27.304	25.476 % ± 13.011	24.028 % ± 13.011
	20(+1)	33.646 % ± 25.411	38.369 % ± 29.360	24.217 % ± 13.511	24.322 % ± 13.511
	25(+1)	–	–	–	–
KS (BW)	15(+2)	32.035 % ± 25.376	43.738 % ± 32.668	26.106 % ± 12.330	31.380 % ± 12.330
	20(+2)	32.002 % ± 25.748	44.153 % ± 33.087	25.198 % ± 12.561	30.624 % ± 12.561
	25(+2)	30.621 % ± 25.646	42.969 % ± 33.173	26.938 % ± 12.406	33.560 % ± 12.406

Tabla 5.8: Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /i/ (1 de 2).

Error relativo en características aerodinámicas - Vocal: /i/					
Método	n_x	Características Aerodinámicas			
		F0	OQ	SQ	NAQ
IBIF		0.772 % ± 1.089	10.374 % ± 8.588	18.067 % ± 10.657	6.452 % ± 10.657
KF	350	0.976 % ± 1.143	11.653 % ± 8.809	17.199 % ± 10.822	7.349 % ± 10.822
KS (WT)	15(+1)	0.728 % ± 1.044	12.083 % ± 7.580	23.321 % ± 12.037	6.468 % ± 12.037
	20(+1)	0.516 % ± 0.770	11.686 % ± 7.228	22.555 % ± 11.461	6.375 % ± 11.461
	25(+1)	0.636 % ± 0.690	11.181 % ± 7.050	21.341 % ± 11.239	6.582 % ± 11.239
KS (RW)	15(+1)	0.831 % ± 1.069	11.595 % ± 7.831	23.494 % ± 11.340	6.465 % ± 11.340
	20(+1)	0.966 % ± 1.379	11.580 % ± 7.427	23.378 % ± 11.212	7.096 % ± 11.212
	25(+1)	–	–	–	–
KS (BW)	15(+2)	0.751 % ± 0.768	12.136 % ± 7.199	31.583 % ± 16.398	8.796 % ± 16.398
	20(+2)	0.724 % ± 0.740	12.017 % ± 7.007	30.803 % ± 16.533	8.849 % ± 16.533
	25(+2)	0.580 % ± 0.613	11.871 % ± 6.885	30.219 % ± 16.435	8.980 % ± 16.435

Tabla 5.9: Error relativo promedio (junto a su desviación estándar) al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada sujeto, considerando la vocal /i/ (2 de 2).

el de comparar las características aerodinámicas obtenidas en cada ventana de tiempo por separado, pues en el otro caso podría ocurrir que existan errores que se compensen, dando como resultado un error menor al real. Para llevar a cabo este análisis, se realizan dos comparaciones, la primera está relacionada a comparar los errores relativos ventana a ventana entre todos los sujetos, mientras que la segunda está relacionada al análisis de dependencia lineal que existe entre las características estimadas con las de referencia (similar a lo que se hace en [22]).

A continuación, en las Tablas 5.10 y 5.11 se muestran los errores relativos promedio cometidos al comparar las características aerodinámicas para cada ventana de tiempo, considerando la vocal /a/ e /i/, respectivamente. En este caso, podemos ver, por ejemplo, que el valor del error relativo en ACFL para el caso de IBIF aumenta, al igual que en el caso de KS - WT, lo que confirma que en las Tablas 5.6, 5.7, 5.8 y 5.9 los errores se compensaban. En el caso del enfoque KS - WT, para la /a/ éste entrega menores errores relativos en comparación con IBIF en 4 casos: H1-H2, F0, SQ y NAQ, mientras que en el caso de la /i/ son 7 casos: ACFL, MFDR, H1-H2, HRF, F0, OQ y NAQ. Por otro lado, en el caso del enfoque KS - BW, para la /a/ éste entrega menores errores relativos en comparación con IBIF en 4 casos: MFDR, H1-H2, HRF y NAQ, mientras que en el caso de la /i/ son 5 casos: ACFL, MFDR, H1-H2, F0 y OQ.

En el caso del análisis de dependencia lineal que existe entre las características estimadas con las de referencia, éste se realiza mediante la regresión lineal entre las características estimadas y las de referencia. Por ejemplo, en la Fig. 5.3 se grafican los valores de ACFL obtenidos para cada ventana de tiempo, para todos los sujetos, considerando el valor de referencia como variable dependiente y considerando el valor estimado mediante el enfoque de KS - WT de orden 20 como variable independiente. Además de estos puntos, se grafica la línea 1:1 que representa una relación lineal perfecta y también se muestran la ecuación lineal que relaciona ambas características junto al valor R^2 . El caso ideal sería uno donde la pendiente fuese exactamente 1, el intercepto fuese 0 y $R^2 = 1$. Para realizar una comparación, en las Tablas 5.12 y 5.13 se muestran los valores de R^2 obtenidos en el caso de la vocal /a/ y en el caso de la vocal /i/, respectivamente. En la Tabla de la /a/ tenemos que el enfoque KS posee valores R^2 más cercanos a 1 que el enfoque IBIF, mientras que en el caso de la /i/ existe una mayor igualdad pero con una leve ventaja para el caso de enfoque KS.

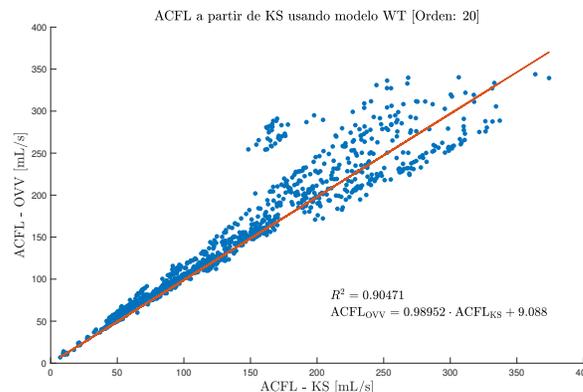


Figura 5.3: Gráfica de ACFL obtenido mediante la señal de referencia y mediante suavizamiento de Kalman usando el modelo de ruido blanco. También se muestra la línea 1:1 para mostrar la dependencia lineal entre el ACFL de referencia con el ACFL estimado.

Error relativo en características aerodinámicas (por ventana) - Vocal: /a/				
Característica	IBIF	KF	KS - WT	KS - BW
		$n_x = 350$	$n_x = 20$	$n_x = 20$
ACFL	7.816 % ± 8.122	8.184 % ± 7.988	8.351 % ± 7.685	8.910 % ± 7.685
MFDR	13.083 % ± 7.895	14.379 % ± 7.692	15.103 % ± 8.843	12.853 % ± 8.843
H1-H2	24.191 % ± 18.616	25.739 % ± 18.383	24.014 % ± 19.333	23.519 % ± 19.333
HRF	26.218 % ± 22.104	27.008 % ± 22.949	26.809 % ± 22.600	24.641 % ± 22.600
F0	0.221 % ± 3.207	0.194 % ± 3.111	0.183 % ± 2.961	0.316 % ± 2.961
OQ	11.135 % ± 8.562	11.441 % ± 8.998	14.488 % ± 10.504	13.701 % ± 10.504
SQ	18.393 % ± 13.358	19.922 % ± 13.355	18.160 % ± 14.272	19.306 % ± 14.272
NAQ	14.193 % ± 8.628	13.402 % ± 8.785	11.338 % ± 7.657	6.821 % ± 7.657

Tabla 5.10: Error relativo promedio al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada ventana de tiempo, considerando la vocal /a/.

Error relativo en características aerodinámicas (por ventana) - Vocal: /i/				
Característica	IBIF	KF	KS - WT	KS - BW
		$n_x = 350$	$n_x = 20$	$n_x = 20$
ACFL	40.312 % ± 24.455	38.986 % ± 24.440	35.361 % ± 23.751	34.083 % ± 23.751
MFDR	43.320 % ± 23.929	42.390 % ± 23.506	38.478 % ± 24.016	41.251 % ± 24.016
H1-H2	33.819 % ± 16.240	35.201 % ± 16.455	33.058 % ± 15.977	33.116 % ± 15.977
HRF	35.211 % ± 16.024	37.442 % ± 16.460	35.120 % ± 15.773	41.697 % ± 15.773
F0	0.476 % ± 4.656	0.432 % ± 4.411	0.323 % ± 3.764	0.450 % ± 3.764
OQ	14.574 % ± 11.412	14.549 % ± 11.491	13.480 % ± 12.214	13.642 % ± 12.214
SQ	22.194 % ± 17.452	22.355 % ± 17.813	25.166 % ± 19.382	31.432 % ± 19.382
NAQ	8.070 % ± 6.290	8.168 % ± 6.163	8.051 % ± 6.341	10.229 % ± 6.341

Tabla 5.11: Error relativo promedio al comparar las características aerodinámicas obtenidas mediante el flujo de aire glotal estimado con la señal de referencia, para cada ventana de tiempo, considerando la vocal /i/.

R^2 - Vocal: /a/				
Característica	IBIF	KF $n_x = 350$	KS - WT $n_x = 20$	KS - BW $n_x = 20$
ACFL	0.8983	0.8956	0.9047	0.9111
MFDR	0.9250	0.9233	0.9253	0.9302
H1-H2	0.3298	0.3634	0.3519	0.3605
HRF	0.4377	0.4329	0.4210	0.5541
F0	0.8109	0.8281	0.8542	0.7719
OQ	0.3602	0.3784	0.3070	0.3235
SQ	0.3124	0.3243	0.3607	0.3412
NAQ	0.6999	0.6675	0.7214	0.7819

Tabla 5.12: Valores R^2 obtenidos al realizar una regresión lineal entre las características estimadas y las características de referencia, para cada ventana de tiempo, considerando la vocal /a/.

R^2 - Vocal: /i/				
Característica	IBIF	KF $n_x = 350$	KS - WT $n_x = 20$	KS - BW $n_x = 20$
ACFL	0.8388	0.8428	0.8301	0.8107
MFDR	0.7818	0.7854	0.7660	0.7229
H1-H2	0.1262	0.1537	0.2220	0.2114
HRF	0.3044	0.3184	0.3491	0.3296
F0	0.4531	0.5138	0.6627	0.5083
OQ	0.0797	0.3215	0.3652	0.3583
SQ	0.3448	0.3550	0.3437	0.3249
NAQ	0.6539	0.6692	0.6842	0.7006

Tabla 5.13: Valores R^2 obtenidos al realizar una regresión lineal entre las características estimadas y las características de referencia, para cada ventana de tiempo, considerando la vocal /i/.

5.4. Tiempos de cómputo

Finalmente, un factor importante a considerar es la complejidad computacional de cada esquema. Aquí es importante notar que el esquema de suavizamiento de Kalman es comparado principalmente con el enfoque de filtraje de Kalman [23], debido a que ambos son esquemas Bayesianos que se implementan recursivamente, a diferencia de IBIF.

Una primera cuestión a considerar es la diferencia entre los órdenes del modelo en variables de estado utilizado en cada enfoque. En el esquema de filtraje de Kalman propuesto en [23], el orden del modelo utilizado es 350, mientras que el enfoque basado en un suavizador de Kalman propuesto en esta tesis considera modelos de orden 20. Esto supone una reducción importante en la complejidad computacional, debido a que el orden del modelo en variables de estado utilizado define los tamaños de las matrices y vectores con los cuales se trabajará. Por ejemplo, tener un modelo de orden n implica trabajar con matrices de orden n^2 . Esto claramente tendrá un impacto en el tiempo que se demore el algoritmo en obtener las estimaciones, independiente de cual sea su implementación. Para cuantificar lo dicha anteriormente, en la Tabla 5.14 se muestran distintas estadísticas con respecto al tiempo necesario para obtener estimaciones del flujo de aire glotal, considerando todos los sujetos y todas la vocales. Para tener como referencia, el mínimo de muestras procesadas fueron 97 296, equivalentes a 11.877 segundos de grabación, mientras que el máximo de muestras fueron 162 615, equivalentes a 19.851 segundos de grabación. En la tabla es posible ver la gran diferencia de tiempos entre los enfoques KF y KS. En el peor de los casos para el enfoque KS, el procesamiento durará 6.8336 segundos, mientras que en el mejor de los casos para el enfoque KF, el procesamiento durará 247.8125 segundos, lo que es 36 veces superior. Si ahora realizamos la comparación con respecto a IBIF, éste último claramente posee menores tiempos de cómputo, sin embargo, gracias a la disminución de orden en el modelo en variables de estado propuesto en esta tesis para el suavizador de Kalman, ahora un enfoque Bayesiano es más competitivo en cuanto al tiempo de cómputo que alternativas Bayesianas anteriores.

Adicionalmente, el menor orden del esquema KS en comparación con el esquema KF, hace que ahora sea mucho más viable implementar la solución al problema de estimación Bayesiana en un dispositivo *wearable* con una capacidad computacional limitada. Lo anterior, es debido que en el caso del esquema KF, es necesario mantener en memoria matrices de órdenes 350×350 (como es el caso de la matriz P_t). En comparación, el esquema KS de orden 20, soluciona este problema al disminuir una matriz de orden 350×350 a una matriz de orden 20×20 .

Tiempo de cómputo (en segundos)			
Estadísticas	IBIF	KF	KS
		$n_x = 350$	$n_x = 20$
Mínimo	0.0144	247.9125	3.6417
Promedio	0.0171	298.6864	4.5747
Máximo	0.0222	418.9291	6.8336

Tabla 5.14: Tiempos de cómputo necesarios para obtener estimaciones del flujo de aire glotal considerando diferentes enfoques.

CONCLUSIONES

En la presente tesis se ha estudiado un enfoque alternativo para solucionar el problema de estimación del flujo de aire glotal a partir de mediciones ambulatorias y no invasivas obtenidas de un acelerómetro puesto sobre la piel del cuello. Se ha propuesto una solución al problema de filtraje inverso mediante una perspectiva Bayesiana, específicamente mediante el esquema de suavizamiento de Kalman.

Para realizar lo anterior, el primer objetivo era obtener una representación en variables de estado de bajo orden a partir del modelo basado en impedancia (IB) subglotal, el cual originalmente fue formulado en términos de su respuesta en frecuencia. Así, se consideraron tres métodos distintos para obtener dicho modelo en variables de estado: reducción de orden, identificación en el dominio del tiempo tiempo e identificación en el dominio de la frecuencia. En el primer caso, se consideró un modelo en variables de estado inicial de alto orden obtenido a partir de la respuesta a impulso del sistema asociada al modelo IB subglotal, para luego aplicar una reducción de modelo mediante truncamiento balanceado. Por otro lado, en el caso de la identificación de sistemas, se usaron datos en el dominio del tiempo, representados por la respuesta a impulso, y datos en el dominio de la frecuencia, representados por la respuesta en frecuencia. En ambos casos, un modelo inicial es estimado mediante métodos de subespacios, para luego aplicar métodos de predicción del error (PEM).

En los procedimientos utilizados se debe elegir como parámetro de entrada el orden del sistema. Así, para poder decidir con qué método obtener la representación en variables de estado y qué orden es el más apropiado, se comparó la raíz del error cuadrático medio que resulta al replicar las respuestas en frecuencia y a impulso del modelo IB subglotal original. Los análisis numéricos realizados muestran que el método que mejor resultados entregó fue el de identificación en el dominio de la frecuencia. Lo anterior tiene sentido, considerando que el modelo IB original se construye numérica y directamente en el dominio de la frecuencia. En los otros dos casos (reducción por truncamiento balanceado e identificación en el dominio del tiempo) se utilizó la respuesta a impulso causal estimada a partir de la respuesta en frecuencia del modelo original. Esta transformación de los datos originales (del dominio de la frecuencia al dominio del tiempo), requirió forzarla a ser causal, lo que resultaba en una posible distorsión de la información en el modelo original. Respecto a la elección de orden, se analizaron 20 órdenes distintos, entre 5 y 100, con un paso de 5, eligiéndose finalmente

como potenciales candidatos los órdenes 15, 20, 25, 30 y 35. Esta es la primera contribución de la presente tesis: obtener una representación en variables de estado del modelo basado en impedancia subglotal de bajo orden. De hecho, el orden de los modelos estimados es aproximadamente 10 veces menor que enfoques Bayesianos anteriores, al mismo tiempo de que se lograron replicar adecuadamente las respuestas en frecuencia y a impulso.

Una vez obtenido el modelo en variables de estado, el siguiente paso era estimar el flujo de aire glotal a partir de un suavizador de Kalman. Para ello, fue necesario introducir la entrada del sistema, que era la señal a estimar, dentro del estado del sistema, que es lo que realmente se estima. Esto se realizó mediante el uso de tres modelos distintos para describir la entrada y así introducirla en el estado: el modelo de ruido blanco, el modelo *random walk* y el modelo de Butterworth. Las estimaciones obtenidas fueron buenas y comparables con enfoques anteriores. Estos resultados constituyen la principal contribución de esta tesis, la cual trata sobre la estimación del flujo de aire glotal mediante un suavizador de Kalman de bajo orden. En específico, en la presente tesis se propuso un esquema alternativo para la realización del filtraje inverso en el contexto de estimación Bayesiana, que permite obtener resultados comparables a enfoques anteriores pero a mucho menor costo computacional debido al bajo orden de los modelos de estado utilizados.

Una de las principales ventajas de los enfoques Bayesianos con respecto al enfoque IBIF, es la de entregar una medida sobre la incertidumbre en las estimaciones, la cual viene dada por la varianza del error de estimación. En un contexto ambulatorio, donde no se puede hacer más que suponer que el modelo utilizado es el modelo verdadero, la medida anterior resulta ser información valiosa, debido a que nos permite tener una noción del error que, según el modelo que estamos utilizando, podríamos estar cometiendo. En esta perspectiva, la varianza del error de estimación obtenida con el suavizador de Kalman de bajo orden propuesto en esta tesis, muestra resultados similares a los obtenidos por otros enfoques Bayesianos en la literatura, pero a un menor costo computacional.

Finalmente, para comparar los resultados de forma numérica, se obtienen diferentes métricas que tienen como objetivo comparar las estimaciones entregadas por los distintos enfoques. La primera métrica es la raíz del error cuadrático medio obtenido al comparar la señal de referencia con la señal estimada. En este caso, el enfoque de suavizamiento de Kalman tuvo mejores resultados que los enfoques anteriores, tanto para la vocal /a/ como para la vocal /i/. En específico, la mayor mejora se logró para la vocal /i/, lo que nos habla de una segunda ventaja del enfoque Bayesiano con respecto a otros esquemas y que guarda relación con el hecho de que el esquema de suavizamiento de Kalman considera que las estimaciones y el modelo poseen errores o, dicho de forma probabilística, poseen ruido. Además, una tercera ventaja es el hecho de que el suavizador de Kalman necesita de parámetros, estos son, las varianzas de los ruidos, los cuales pueden ser utilizados como parámetros de diseño de tal manera de mejorar las estimaciones.

La segunda comparación que se realizó, fue en el contexto de estimar las características aerodinámicas del flujo de aire glotal. Para ello, se utilizó en primera instancia el error relativo al comparar cada característica aerodinámica obtenida de promediar sus valores en todas las ventanas de 50 ms. Los mejores resultados del enfoque de suavizamiento de Kalman fueron los que utilizaron el modelo de ruido blanco y de Butterworth. En ambos casos se obtiene, en resumen, resultados comparativos a los entregados por el enfoque de IBIF. En algunas características el enfoque KS entregó menores errores relativos con respecto a IBIF, mientras que en otras entregó mayores errores. En específico, el enfoque de suavizamiento de Kalman que utiliza el modelo de ruido blanco, entregó menores errores relativos que IBIF en 10 características (considerando la vocal /a/ y la vocal /i/) y en sólo 6 fueron mayores.

6.1. Trabajo futuro

Es evidente que en un trabajo de investigación se van abriendo nuevos caminos a medida que se avanza. Como es imposible incluir todas las posibles exploraciones, a continuación se describen algunos caminos que, según el autor, podrían seguirse con el fin de mejorar el enfoque aquí propuesto:

- Una ventaja que posee el enfoque de suavizamiento de Kalman es que es posible agregar un modelo del flujo de aire glotal, con el fin de darle cierta *forma* a las estimaciones y así mejorarlas. En este trabajo sólo se consideró el modelo de Butterworth, el cual es un modelo simple pero efectivo que mostró buenos resultados. Sin embargo, existen más modelos glotales, los cuales podrían ser estudiados en este contexto. Por lo anterior, una primera incursión que podría mejorar sustancialmente las estimaciones, es la de agregar otros modelos glotales para el modelado de la entrada u_t^g . Posibles modelos serían el de KLGLOTT88, Rosenberg C, R++ y LF. Éste último fue propuesto para describir la derivada del flujo de aire glotal, por lo que sería de especial interés utilizarlo con el fin de obtener buenas estimaciones en la característica aerodinámica MFDR, la cual cuantifica la velocidad de cierre de las cuerdas vocales.
- El suavizador de Kalman es un esquema que entrega estimaciones *offline*, esto es, se requiere de una ventana de datos para estimar la señal dentro de ella. En este trabajo sólo se analizó la estimación del flujo de aire glotal utilizando todo el conjunto de datos disponibles. Por lo anterior, sería interesante estudiar cómo implementar este esquema de la manera más *online* posible, esto es, se podría analizar el ancho de la ventana de datos óptima de tal manera de no comprometer la precisión, al mismo tiempo de no necesitar un conjunto de datos muy grande que haga que, por ejemplo, las estimaciones sean entregadas con un gran retraso.
- Al obtener la respuesta a impulso del modelo IB subglotal, hemos dicho que ésta era no causal. A pesar de lo anterior, el modelo representa un sistema físico, por lo que dicha respuesta debería ser causal. Así, sería interesante responder por qué $h_t \neq 0$ para $t < 0$. Un fenómeno que podría ser considerado es el de *aliasing* en el tiempo.
- Por simplicidad, en este trabajo se supuso que los ruidos de proceso sólo excitaban un estado (el que se quería estimar). Una pregunta que queda abierta es cuál sería el impacto en las estimaciones si se considerara ruido de proceso en todos los estados del modelo, esto es, que la diagonal completa de la matriz Q sea distinta de cero.
- Para un sistema lineal e invariante en el tiempo, la ganancia del filtro/suavizador de Kalman converge, al igual que la varianza del error de estimación. Lo anterior plantea la pregunta de qué tanto empeorarían las estimaciones si ahora se considerase un suavizador de Kalman estacionario. Este último ya no funcionaría de manera iterativa, lo que supondría una mejora en cuanto a los tiempos de cómputo (quizá incluso comparables con los de IBIF).
- Como se vio en el capítulo 5, el RMSE no sirve del todo para predecir cuándo un modelo es bueno o malo para estimar las características aerodinámicas. Lo anterior es debido a que existían casos donde el RMSE era bajo pero que no se veía reflejado al momento de estimar las características aerodinámicas. Un posible trabajo futuro sería el de buscar una métrica que permita elegir efectivamente qué modelo y parámetros (σ_w^2, σ_v^2) elegir, con el fin de minimizar el error cometido al estimar dichas características.

DISTRIBUCIÓN GAUSSIANA

Definición A.1 (Distribución Gaussiana [66]). Una variable aleatoria $\mathbf{x} \in \mathbb{R}^n$ tiene distribución Gaussiana con media $\boldsymbol{\mu} \in \mathbb{R}^n$ y covarianza $\mathbf{P} \in \mathbb{R}^{n \times n}$ si su función de densidad tiene la forma

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{P}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (\text{A.1})$$

donde $|\mathbf{P}|$ es el determinante de la matriz \mathbf{P} . Lo anterior se denota como $\mathbf{x} \sim \mathcal{N}_{\mathbf{x}}(\boldsymbol{\mu}, \mathbf{P})$.

Lema A.1 (Distribución conjunta de variables Gaussianas [47]). Si las variables aleatorias $\mathbf{x} \in \mathbb{R}^n$ y $\mathbf{y} \in \mathbb{R}^m$ tienen funciones de probabilidad Gaussianas

$$\mathbf{x} \sim \mathcal{N}_{\mathbf{x}}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{P}_{\mathbf{x}}) \quad (\text{A.2})$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}_{\mathbf{y}}(\mathbf{C}\mathbf{x} + \boldsymbol{\mu}, \mathbf{P}_{\mathbf{y}|\mathbf{x}}) \quad (\text{A.3})$$

entonces la distribución conjunta de \mathbf{x} , \mathbf{y} y la distribución marginal de \mathbf{y} están dadas por

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}_{(\mathbf{x}, \mathbf{y})}^\top \left(\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \mathbf{C}\boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{\mathbf{x}} & \mathbf{P}_{\mathbf{x}}\mathbf{C}^\top \\ \mathbf{C}\mathbf{P}_{\mathbf{x}} & \mathbf{C}\mathbf{P}_{\mathbf{x}}\mathbf{C}^\top + \mathbf{P}_{\mathbf{y}|\mathbf{x}} \end{pmatrix} \right) \quad (\text{A.4})$$

$$\mathbf{y} \sim \mathcal{N}_{\mathbf{y}}(\mathbf{C}\boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}, \mathbf{C}\mathbf{P}_{\mathbf{x}}\mathbf{C}^\top + \mathbf{P}_{\mathbf{y}|\mathbf{x}}) \quad (\text{A.5})$$

Demostración. Ver [47]. ■

Lema A.2 (Distribución condicional de variables Gaussianas [47]). Si las variables aleatorias $\mathbf{x} \in \mathbb{R}^n$ y $\mathbf{y} \in \mathbb{R}^m$ tienen la función de probabilidad conjunta Gaussiana

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}_{(\mathbf{x}, \mathbf{y})}^\top \left(\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{\mathbf{x}} & \mathbf{P}_{\mathbf{x}\mathbf{y}} \\ \mathbf{P}_{\mathbf{x}\mathbf{y}}^\top & \mathbf{P}_{\mathbf{y}} \end{pmatrix} \right) \quad (\text{A.6})$$

entonces la distribución marginal y condicional de \mathbf{x} y \mathbf{y} están dadas por

$$\mathbf{x} \sim \mathcal{N}_{\mathbf{x}}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{P}_{\mathbf{x}}) \quad (\text{A.7})$$

$$\mathbf{y} \sim \mathcal{N}_y(\boldsymbol{\mu}_y, \mathbf{P}_y) \quad (\text{A.8})$$

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}_x(\boldsymbol{\mu}_x + \mathbf{P}_{xy}\mathbf{P}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \mathbf{P}_x - \mathbf{P}_{xy}\mathbf{P}_y^{-1}\mathbf{P}_{xy}^\top) \quad (\text{A.9})$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}_y(\boldsymbol{\mu}_y + \mathbf{P}_{xy}^\top\mathbf{P}_x^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \mathbf{P}_y - \mathbf{P}_{xy}^\top\mathbf{P}_x^{-1}\mathbf{P}_{xy}) \quad (\text{A.10})$$

Demostración. Ver [47].

■

BIBLIOGRAFÍA

- [1] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, “Voice disorders in the general population: Prevalence, risk factors, and occupational impact,” *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.
- [2] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray, and E. M. Smith, “Prevalence of voice disorders in teachers and the general population,” *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 2, pp. 281–293, 2004.
- [3] N. Bhattacharyya, “The prevalence of voice problems among adults in the United States,” *Laryngoscope*, vol. 124, no. 10, pp. 2359–2362, 2014.
- [4] K. M. Mora Pino, F. J. Clavijo Rocha, S. P. Galdames Durán, C. C. Maya Molina, and V. A. Soto Galleguillos, “Contexto ocupacional, abuso y mal uso vocal en profesores de la ciudad de Iquique,” *Ciencia & trabajo*, vol. 20, no. 62, pp. 116–120, 2018.
- [5] J. N. Figueredo Ruiz and J. A. Castillo Martínez, “Evaluación de desórdenes vocales en profesionales que usan su voz como herramienta de trabajo,” *Revista Ciencias de la Salud*, vol. 14, pp. 97–112, 2015.
- [6] A. Castillo, C. Casanova, D. Valenzuela, and S. Castañón, “Prevalencia de disfonía en profesores de colegios de la comuna de Santiago y factores de riesgo asociados,” *Ciencia & trabajo*, vol. 17, no. 52, pp. 15–21, 2015.
- [7] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, “Objective assessment of vocal hyperfunction,” *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 2, pp. 373–392, 1989.
- [8] R. E. Hillman, C. E. Stepp, J. H. Van Stan, M. Zañartu, and D. D. Mehta, “An updated theoretical framework for vocal hyperfunction,” *American Journal of Speech-Language Pathology*, vol. 29, no. 4, pp. 2254–2260, 2020.
- [9] S. B. Davis, “Acoustic characteristics of normal and pathological voices,” *Speech and Language*, vol. 1, pp. 271–335, 1979.
- [10] D. D. Mehta and R. E. Hillman, “Use of aerodynamic measures in clinical voice assessment,” *Perspectives on Voice and Voice Disorders*, vol. 17, no. 3, 2007.
- [11] E. B. Holmberg, P. Doyle, J. S. Perkell, B. Hammarberg, and R. E. Hillman, “Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation

- in baseline and changes across voice therapy,” *Journal of Voice*, vol. 17, no. 3, pp. 269–282, 2003.
- [12] J. H. Van Stan, D. D. Mehta, S. M. Zeitels, J. A. Burns, A. M. Barbu, and R. E. Hillman, “Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects,” *Annals of Otology, Rhinology and Laryngology*, vol. 124, no. 11, pp. 864–874, 2015.
- [13] V. M. Espinoza, M. Zañartu, J. H. Van Stan, D. D. Mehta, and R. E. Hillman, “Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction,” *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 8, pp. 2159–2169, 2017.
- [14] M. Rothenberg, “A new inverse-filtering technique for deriving the glottal air flow waveform during voicing,” *The Journal of the Acoustical Society of America*, vol. 53, no. 6, pp. 1632–1645, 1973.
- [15] D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne, and R. E. Hillman, “Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 11, pp. 3090–3096, 2012.
- [16] D. D. Mehta, J. H. Van Stan, M. Zañartu, M. Ghassemi, J. V. Guttag, V. M. Espinoza, J. P. Cortés, H. A. Cheyne, and R. E. Hillman, “Using ambulatory voice monitoring to investigate common voice disorders: Research update,” *Frontiers in Bioengineering and Biotechnology*, vol. 3, pp. 1–14, 2015.
- [17] J. P. Cortés, V. M. Espinoza, M. Ghassemi, D. D. Mehta, J. H. Van Stan, R. E. Hillman, J. V. Guttag, and M. Zañartu, “Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration,” *PLoS ONE*, vol. 13, no. 12, pp. 1–22, 2018.
- [18] M. Ghassemi, J. H. Van Stan, D. D. Mehta, M. Zañartu, H. A. Cheyne, R. E. Hillman, and J. V. Guttag, “Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1668–1675, 2014.
- [19] A. V. Oppenheim, A. S. Willsky, and S. Hamid Nawab, *Signals and systems*. Prentice Hall, 1997.
- [20] M. E. Salgado, J. I. Yuz, and R. A. Rojas, *Análisis de sistemas lineales*. Prentice Hall, 2014.
- [21] V. M. Espinoza, D. D. Mehta, J. H. Van Stan, R. E. Hillman, and M. Zañartu, “Glottal aerodynamics estimated from neck-surface vibration in women with phonotraumatic and nonphonotraumatic vocal hyperfunction,” *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 9, pp. 2861–2869, sep 2020.
- [22] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, “Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 9, pp. 1929–1939, 2013.

- [23] J. P. Cortés, “Physiologically based features related to vocal hyperfunction: From clinical to ambulatory data,” Ph.D. dissertation, Universidad Técnica Federico Santa María, 2020.
- [24] L. J. Boë, T. R. Sawallis, J. Fagot, P. Badin, G. Barbier, G. Captier, L. Ménard, J. L. Heim, and J. L. Schwartz, “Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science,” *Science Advances*, vol. 5, no. 12, 2019.
- [25] J. Markel and A. Gray, *Linear prediction of speech*. Springer, 1976.
- [26] L. R. Rabiner and R. W. Schafer, *Theory and application of digital speech processing*. Pearson, 2010, vol. 109, no. 5.
- [27] G. Fant, *Acoustic theory of speech production*. Walter de Gruyter, 1970.
- [28] B. Doval, C. Alessandro, and N. H. Bernardoni, “The spectrum of glottal flow models,” *Acta Acustica united with Acustica*, vol. 92, no. 6, pp. 1026–1046, 2006.
- [29] A. E. Rosenberg, “Effect of glottal pulse shape on the quality of natural vowels,” *The Journal of the Acoustical Society of America*, vol. 49, no. 1A, pp. 583–590, 1971.
- [30] R. C. Dorf and J. A. Svoboda, *Introduction to Electric Circuits*, 9th ed. John Wiley & Sons, 2001.
- [31] K. J. Åström and B. Wittenmark, *Computer Controlled Systems*. Prentice Hall, 1997.
- [32] V. Espinoza, “Stationary And Dynamic Aerodynamic Assessment Of Vocal Hyperfunction Using Enhanced Supraglottal And Subglottal Inverse Filtering Methods,” Ph.D. dissertation, Universidad Técnica Federico Santa María, 2018.
- [33] L. Ljung, *System identification: Theory for the user*. Prentice Hall, 1999.
- [34] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 2001.
- [35] T. McKelvey, “Identification of State-Space Models from Time and Frequency Data,” Ph.D. dissertation, Linköping University, 1995.
- [36] B. L. Ho and R. E. Kalman, “Effective construction of linear state-variable models from input/output functions,” *At-Automatisierungstechnik*, vol. 14, no. 1-12, pp. 545–548, 1966.
- [37] H. Akaike, “Stochastic Theory of Minimal Realization,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 667–674, 1974.
- [38] H. Zeiger and McEwen, “Approximate linear realizations of given dimension via Ho’s algorithm,” *IEEE Transactions on Automatic Control*, vol. 19, no. 2, pp. 153 – 153, 1974.
- [39] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996.
- [40] T. Katayama, *Subspace Methods for System Identification*. Springer, 2004.
- [41] M. Viberg, “Subspace-based methods for the identification of linear time-invariant systems,” *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.
- [42] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, 1983.

- [43] R. L. Burden and J. D. Faires, *Numerical Analysis*. Brooks/Cole, 2011.
- [44] K. Zhou, J. Doyle, and K. Glover, *Robust and optimal control*. Prentice-Hall, 1996.
- [45] L. Pernebo and L. M. Silverman, “Model Reduction via Balanced State Space Representations,” *IEEE Transactions on Automatic Control*, vol. 27, no. 2, pp. 382–387, 1982.
- [46] A. J. Laub, M. T. Heath, C. C. Paige, and R. C. Ward, “Computation of System Balancing Transformations and Other Applications of Simultaneous Diagonalization Algorithms,” *IEEE Transactions on Automatic Control*, vol. 32, no. 2, pp. 115–122, 1987.
- [47] S. Sarkka, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.
- [48] B. D. O. Anderson and J. B. Moore, *Optimal filtering*. Prentice-Hall, 1979.
- [49] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*. John Wiley & Sons, 1950.
- [50] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960.
- [51] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” *Journal of Fluids Engineering, Transactions of the ASME*, vol. 83, no. 1, pp. 95–108, 1961.
- [52] S. Haykin, *Kalman filtering and neural networks*. John Wiley & Sons, 2001.
- [53] A. H. Jazwinski, *Stochastic processes and filtering theory*, ser. Mathematics in science and engineering. Academic Press, 1970.
- [54] T. Söderström, *Discrete-time stochastic systems*. Springer, 2002.
- [55] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965.
- [56] L. Ljung and T. Glad, *Modeling of Dynamic Systems*. Prentice Hall, 1994.
- [57] K. Ishizaka, M. Matsudaira, and T. Kaneko, “Input acoustic-impedance measurement of the subglottal system,” *The Journal of the Acoustical Society of America*, no. 1, pp. 190–197, 1976.
- [58] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, “An acoustic model of the respiratory tract,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 543–550, 2001.
- [59] J. C. Ho, M. Zañartu, and G. R. Wodicka, “An anatomically based, time-domain acoustic model of the subglottal system for speech production,” *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1531–1547, 2011.
- [60] K. Ishizaka, J. C. French, and J. L. Flanagan, “Direct Determination of Vocal Tract Wall Impedance,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 4, pp. 370–373, 1975.
- [61] G. Degottex, “Glottal source and vocal-tract separation,” Ph.D. dissertation, Université Pierre et Marie Curie - Paris VI, 2011.

- [62] M. Rothenberg, “The Glottal Volume Velocity Waveform during Loose and Tight Voiced Glottal Adjustments,” in *7th International Congress of Phonetic Sciences*, 1971, pp. 380–388.
- [63] A. V. Oppenheim and R. W. Schaffer, *Discrete-time signal processing*. Pearson, 2010.
- [64] J. Glover, “The linear estimation of completely unknown signals,” *IEEE Transactions on Automatic Control*, vol. 14, no. 6, pp. 766 – 767, 1969.
- [65] M. Verhaegen and V. Verdult, *Filtering and system identification: A least squares approach*. Cambridge University Press, 2007, vol. 9780521875.
- [66] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes*, 4th ed. McGraw-Hill, 2002.