

**UCLV**  
Universidad Central  
"Marta Abreu" de Las Villas



**FIE**  
Facultad de  
Ingeniería Eléctrica

Departamento de Telecomunicaciones y  
Electrónica

## TRABAJO DE DIPLOMA

“Relación entre la duración del pulso glotal y sus fases en señales EGG”

**Autor:** Reinier Rodríguez Guillén

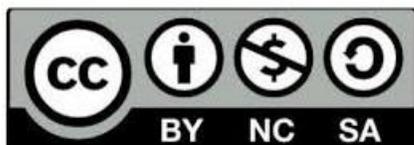
**Tutor:** Dr. Carlos Ariel Ferrer Riesgo

Santa Clara, Junio de 2019  
Copyright©UCLV

Este documento es Propiedad Patrimonial de la Universidad Central “Marta Abreu” de Las Villas, y se encuentra depositado en los fondos de la Biblioteca Universitaria “Chiqui Gómez Lubian” subordinada a la Dirección de Información Científico Técnica de la mencionada casa de altos estudios.

Se autoriza su utilización bajo la licencia siguiente:

**Atribución- No Comercial- Compartir Igual**



Para cualquier información contacte con:

Dirección de Información Científico Técnica. Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní. Km 5½. Santa Clara. Villa Clara. Cuba. CP. 54 830

Teléfonos.: +53 01 42281503-1419



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería en Telecomunicaciones y Electrónica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

---

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

---

Firma del Tutor

---

Firma del Jefe de Departamento  
donde se defiende el trabajo

---

Firma del Responsable de  
Información Científico-Técnica

## **PENSAMIENTO**

**“ESCUCHAMOS LO QUE NADIE HA DICHO; MIRAMOS LO QUE  
NADIE VE”.**

**KAKUZOO KAKURA**

## **DEDICATORIA**

Dedico este trabajo a mi mamá la cual se ha sacrificado mucho por mí y lo sigue haciendo, a mi papá y a mi abuela Marta que siempre han estado presentes. A mi abuelo fallecido Papi que siempre me apoyó. Una dedicatoria especial a mi niña Emily y a mi novia Yainet, por ser las personas que más quiero. A mi familia.

## **AGRADECIMIENTOS**

Son muchas personas las personas a las que agradecerles. Si algunas no están, que me disculpen, es por mi falta de memoria. A mi esposa y bebe que me han impulsado a terminar la tesis. A mi tutor que me ha ayudado enormemente todo este tiempo, y sin el cual no habría podido terminarla. A mi mamá, papá, abuela Marta, tío Sergio, en fin, a mi familia, a mis suegros y compañeros de trabajo.

## TAREA TÉCNICA

Para lograr la confección del presente trabajo, dar cumplimiento a los objetivos trazados y obtener los resultados esperados, se desarrollan las siguientes tareas técnicas:

1. Revisión bibliográfica de los modelos de síntesis del pulso glotal y sus fases.
2. Caracterización de los modelos de síntesis del pulso glotal.
3. Caracterización de las señales EGG.
4. Relación entre las medidas de pulso glotal y sus fases.
5. Evaluación de las medidas de pulso glotal y sus fases.
6. Elaboración del informe final del Trabajo de Diploma.

---

Firma del Autor

---

Firma del Tutor

## RESUMEN

El habla está formada por el flujo de aire que viene desde los pulmones y pasa a través de la glotis, las cuerdas vocales, el tracto vocal, la lengua, las mejillas y finalmente los labios. Una forma que han encontrado los ingenieros para entender este fenómeno ha sido el desarrollo de aplicaciones de voz como decodificadores o sintetizadores de voz. A menudo se usan modelos que generan formas de ondas sintéticas del pulso glotal, que son usadas en sintetizadores de voz de acuerdo al modelo fuente-filtro. Existen diferentes maneras de modificar las ondas sintéticas cuando se considera variable la longitud del pulso(*jitter*). Las voces patológicas muestran una cantidad significativa de *jitter*, para medir la severidad de sus síntomas, se emplean las medidas de perturbación de voz. Los métodos para medir perturbaciones de voz son evaluados en señales sintéticas, por lo que es necesario, que el modelo para generar estas señales sea lo más cercano posible a las voces patológicas reales. En este trabajo se estudia la relación entre la variación de la longitud del pulso glotal y sus fases; se analizan los distintos modelos de síntesis y las características de la señal EGG. Se llega a la conclusión de que el modelo basado en PW con Op constante emerge como el más adecuado a ser usado en los sintetizadores de voz.

## TABLA DE CONTENIDOS

PENSAMIENTO .....	i
DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
TAREA TÉCNICA .....	iv
RESUMEN .....	v
INTRODUCCIÓN .....	1
CAPÍTULO 1. Descripción y análisis de la señal Electroglotográfica y sus fases .....	4
1.1 Modelo de síntesis del pulso glotal. Fuente-Filtro .....	4
1.1.1 Modelo de la Fuente .....	5
1.1.1.1 Modelo Fisiológico .....	6
1.1.1.2 Modelos Matemáticos .....	7
1.1.1.3 Modelos Físicos .....	9
1.1.2 Modelos del Tracto Vocal (Filtro) .....	12
1.2 Perturbaciones Vocales .....	14
1.3 Influencias, Enfoques y Contradicciones de los modelos de inducción de <i>jitter</i> ... 15	15
1.3.1 Modelo <i>Constant-Warping</i> (CW).....	15
1.3.2 Modelo <i>Piecewise-Warping</i> (PW) .....	16
1.4 Señal Electroglotográfica .....	16
1.4.1 Principios Físicos .....	17
1.4.2.1 Detección de Instante de Cierre Glotal (GCI).....	19
1.4.2.2 Detección de Instante de Apertura Glotal (GOI) .....	22
1.5 Conclusiones Parciales .....	24

CAPÍTULO 2. Correlaciones asumidas entre las fases y la longitud del pulso glotal.....	25
2.1 Señales EGG empleadas .....	25
2.1.1 Base de datos de voz <i>Saarbrücken</i> .....	25
2.1.2 Selección de señales EGG .....	26
2.1.3 Selección de segmentos .....	27
2.2 Métodos usados para la detección de GCI y GOI.....	28
2.3 Modelos de relación considerados .....	31
2.4 Medidas de relación entre pulso glotal y sus fases .....	32
2.4.1 Secuencias extraídas .....	32
2.4.2 Correlaciones extraídas.....	33
2.5 Conclusiones parciales .....	34
CAPÍTULO 3. Evaluación de las relaciones encontradas entre las fases y la longitud del pulso	
.....	35
3.1 Análisis matemático de las correlaciones .....	35
3.2 Fuerza de las perturbaciones .....	36
3.3 Relaciones encontradas .....	37
3.4 Relaciones encontradas basadas en la cantidad de jitter .....	37
3.5 Ocurrencia de Outliers .....	38
CONCLUSIONES Y RECOMENDACIONES .....	39
Conclusiones .....	39
Recomendaciones .....	40
REFERENCIAS BIBLIOGRÁFICAS .....	41
ANEXOS .....	45

## INTRODUCCIÓN

La voz es indiscutiblemente la manera más natural y efectiva de comunicación entre seres humanos. La producción de la voz es un proceso físico ejecutado por los humanos para producir sonidos. El flujo de aire que pasa a través de la glotis, es distorsionado de una manera u otra, lo cual produce una variedad de sonidos que llegan al lenguaje. El estudio de la producción del habla requiere la unión de distintos factores: acústico, fisiológico, fonético y de procesamiento de señales. Existen libros relacionados con la producción de la voz que son referentes indiscutibles (Fant.G, 1970, Flanagan, 1972, L.R. Rabiner, 1978, J.R. Deller, 1993).

Para obtener las señales de voces reales se puede usar la señal electroglotográfica (EGG) (Baken.J.R, 1992, Mora, 2014), esta surge de un método inventado en 1957 (Fabre, 1957) basado en la variación de la impedancia de la glotis. La EGG, permite la exploración del comportamiento laríngeo sin necesidad de realizar un procedimiento invasivo. Es una valoración objetiva que permite evaluar el contacto que se produce en los pliegues vocales durante la fonación. Esta señal presenta dos fases llamadas instante de cierre glotal (GCI) e instante de apertura glotal (GOI). Las voces reales pueden presentar jadeo, aspereza entre otras patologías, por lo que es importante su estudio.

En las señales de voces reales se encuentran perturbaciones vocales como *jitter* y *shimmer*. Estas son perturbaciones pequeñas, aleatorias en el tiempo y que varían en la amplitud y en el período. Las voces patológicas tienen una gran cantidad de *jitter* y existen muchos métodos para medirlos. Estos métodos se evalúan mediante el empleo de señales sintéticas, las cuales son simuladas usando software llamados sintetizadores de voz.

Los sintetizadores de voz permiten la producción de la voz de forma muy semejante a la real para lo cual usan formas de ondas sintéticas del pulso glotal. Estos tienen muchas ventajas

en comparación con dispositivos físicos (hardware). Entre las ventajas que brindan encontramos que estos no necesitan una calibración repetida, son estables y tienen una tasa de señal a ruido alta (SNR) (Klatt, 1980a), así como la habilidad de modelar un amplio rango de cualidades vocales. Además, permiten el desarrollo de importantes aplicaciones como prótesis vocales, análisis de sonidos naturales y el modelado de diferentes tipos de fonación (Bangayan et al., 1997). Para sintetizar señales se utilizan distintos modelos como *Constant-Warping* y *Piecewise-Warping*, estos están relacionados con las fases del pulso glotal y su duración. En la actualidad no se ha realizado un estudio que evalúe la relación que existe entre las fases del pulso glotal y su duración, considerando sus fases de apertura y cierre.

Teniendo en cuenta lo anterior expuesto, se enuncia el siguiente **problema científico**:

¿Cómo encontrar la relación que existe entre las fases del pulso glotal y su duración considerando los puntos iniciales y finales de la señal EGG?

Esta investigación tiene como **objetivo de estudio** la relación de las fases del pulso glotal y su duración en señales EGG y el **campo de estudio** lo constituye las señales EGG.

Para dar cumplimiento al problema de investigación se propone el siguiente **objetivo general**: Evaluar la relación que existe entre las fases del pulso glotal y su duración considerando los puntos iniciales y finales (GOI y GCI) en señales EGG.

Para resolver el problema científico y dar cumplimiento al objetivo general se plantean los siguientes **objetivos específicos**:

- 1- Seleccionar un conjunto de señales EGG apropiadas para la evaluación de los modelos de relación.
- 2- Seleccionar y caracterizar un conjunto de posibles modelos de relación entre la duración de las fases interiores del pulso y su duración total.
- 3- Proponer y evaluar un conjunto de criterios que permitan determinar lo ajustado a la verdad de cada modelo.

De los objetivos específicos propuestos surgen las siguientes **interrogantes científicas**:

- 1- ¿Cuál señal EGG es apropiada para la evaluación de los modelos de relación?
- 2- ¿Cómo seleccionar y caracterizar un conjunto de posibles modelos de relación?

3- ¿Cómo evaluar un conjunto de criterios que permitan determinar lo ajustado a la verdad de cada modelo?

El presente trabajo está estructurado de la siguiente forma: resumen, introducción, capítulos, conclusiones, recomendaciones y referencias bibliográficas.

En el **primer capítulo** se exponen los modelos fisiológicos, matemáticos y físicos de la fuente glotal y del tracto vocal. Se explican los diferentes tipos de perturbaciones vocales. Se caracterizan modelos de relación. Se abordan los principios de la señal Electroglotográfica y sus fases. Por último, se llega a conclusiones parciales.

En el **segundo capítulo** se caracteriza a la base de datos usada. Se explican los métodos usados para detección de las fases GCIs y GOIs. Se describen los modelos de relación considerados y posteriormente las medidas de relación entre el pulso glotal y sus fases. Por último, se llega a conclusiones parciales.

En el **tercer capítulo** se desarrolla un análisis matemático de las correlaciones encontradas, posteriormente se explica la fuerza de las perturbaciones y se enumeran las relaciones encontradas.

## CAPÍTULO 1. Descripción y análisis de la señal Electroglotográfica y sus fases

En la primera sección de este capítulo se explicará el modelo fuente-filtro, derivándose de este los modelos matemáticos, físicos y fisiológicos de la fuente y los modelos del filtro. Se explicarán los diferentes tipos de perturbaciones vocales. Se abordarán los modelos usados en los sintetizadores de voz. Se caracterizará a la señal EGG y posteriormente se explicarán las fases de esta y los métodos para su detección.

### 1.1 Modelo de síntesis del pulso glotal. Fuente-Filtro

Uno de los dos elementos principales en la teoría fuente-filtro usada para la producción de la voz es la forma de onda  $g(t)$  del pulso glotal (Fant.G, 1970). Por lo que el estudio de esta es de vital importancia a la hora de entender el habla y modelarla.

El modelo mostrando en la Figura.1.1, presentado en (G.Fant, 1960), es conocido como modelo Fuente-Filtro. La forma de onda acústica audible  $s(t)$ (Ec.1.1) es usualmente modelada como la convolución en el tiempo de la forma de onda del pulso glotal  $g(t)$  con la respuesta al impulso del tracto vocal  $h(t)$  y la radiación de los labios  $r(t)$ . El  $r(t)$  es asumido como la operación de la derivada aplicada a  $g(t)$ , para que la convolución sea entre  $g'(t)$  y  $h(t)$ .

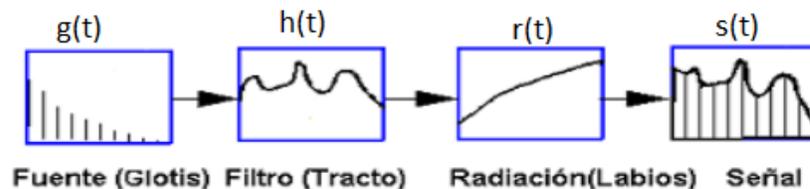
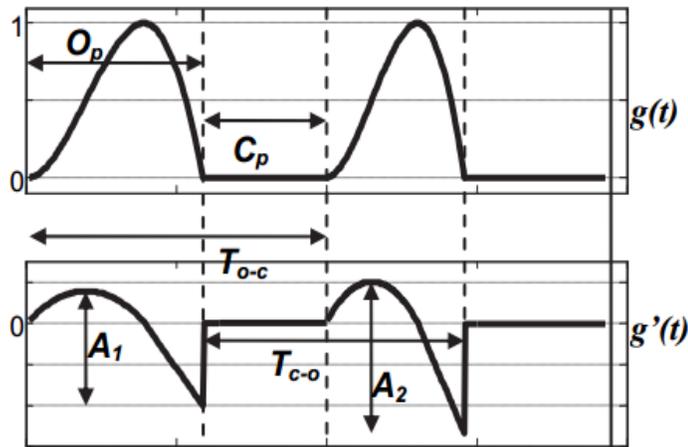


Figura.1.1. Modelo Fuente-Filtro de producción (G.Fant, 1960).

$$s(t) = h(t) * r(t) * \left[ \sum_{i=1}^N a_i * g_i(t) * \delta(t - (T_{i-1} + \bar{T}_0 + \Delta T_{i-1})) \right] + e(t) \quad (1.1)$$

$$s(t) = h(t) * \left[ \sum_{i=1}^N a_i * g'_i(t) * \delta(t - (T_{i-1} + \bar{T}_0 + \Delta T_{i-1})) \right] + e(t)$$

La  $g(t)$  está esquemáticamente dividida en fases de apertura ( $O_p$ ) y cierre ( $C_p$ ), ver (Figura.1.2.).



**Figura.1.2.** Gráfico esquemático de  $g(t)$  y sus fases (superior) y su derivada  $g'(t)$  (inferior) de una señal con jitter (Ferrer et al., 2015).

Durante  $C_p$  no hay un flujo neto de aire (i.e.  $g(t)=0$ ), ya que la glotis se mantiene cerrada. En la derivada  $g'(t)$  (Figura.1.2.(inferior)), el rasgo más relevante es que durante el período de  $O_p$  la señal no se mantiene en cero y que existe un cambio repentino en la transición desde  $O_p$  a  $C_p$  (Ferrer et al., 2015).

### 1.1.1 Modelo de la Fuente

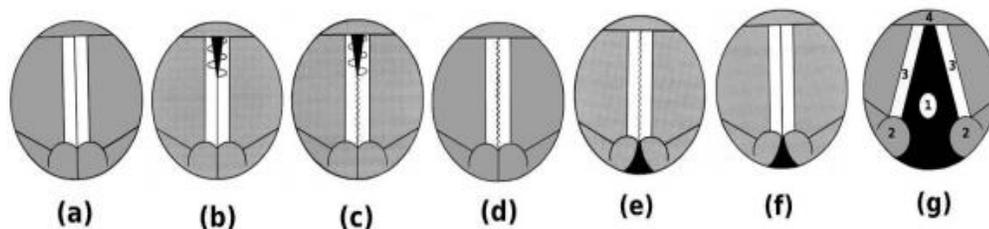
Para simular las características de la fuente glotal, se han desarrollado diversos modelos, estos se pueden dividir en matemáticos, físicos y fisiológicos, los cuales se explicarán en los siguientes epígrafes.

### 1.1.1.1 Modelo Fisiológico

El sistema de producción de la voz humana nos permite producir una gran cantidad de sonidos, este consiste en muchos órganos que cooperan en el proceso de la fonación, el cual está dividido en tres grupos: los pulmones, laringe y el tracto vocal. Desde el punto de vista fisiológico el aire sale de los pulmones y pasa por la laringe donde el flujo de aire es modulado por la vibración de las cuerdas vocales. Las cuerdas vocales vibran convirtiendo el flujo de aire en pulsos acústicos lo cual se convierte en una señal de excitación para el tracto vocal. El tracto vocal consiste en cavidades resonantes nasales, orales y de la faringe que filtran el flujo de aire de la señal modelada. El flujo de aire resultante es luego radiado por los labios. Diferentes tipos de sonidos se pueden lograr modificando los patrones de vibración de las cuerdas vocales, o de la configuración del tracto vocal.

Cuando se habla, la oscilación de las cuerdas vocales interrumpe periódicamente el flujo de aire proveniente de los pulmones y crea cambios en la presión de aire. Cuando ningún sonido es emitido el tracto vocal generalmente está abierto. Para producir sonidos de voz los músculos controlan el cierre de las cuerdas vocales para proveer resistencia al aire que viene de los pulmones. Este ciclo es repetido muchas veces en cada segundo y la duración de este es llamado período fundamental ( $T_0$ ).

Desde el punto de vista de análisis de voz, el procesamiento de la fuente glotal es examinada para ambos estados (modal y no modal). La voz modal generalmente representa al tipo de fonación neutral, la cual tiene una pequeña variación período a período en sucesivos instantes glotales, esta también asume que hay una significativa excitación cerca del instante de cierre de la glotis (GCI). El tipo de fonación no modal involucra significativas variaciones en las características de la fuente glotal. Para estos propósitos diferentes tipos de fonación en habla normal han sido categorizados (Figura.1.3) (Drugman et al., 2014).

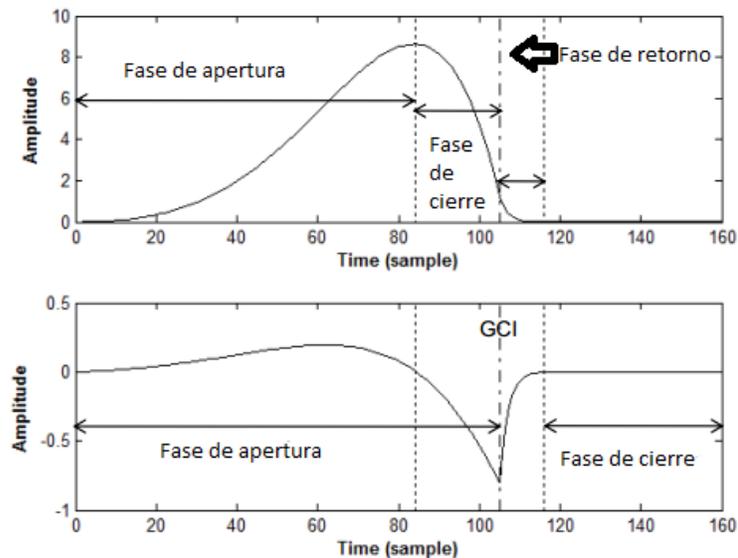


**Figura.1.3. Configuraciones glotales para varios tipos de fonación. (a) Cierre Glotal. (b) Chirrido. (c) Voz chirriante. (d) Voz modal. (e) Voz entrecortada. (f) Susurro. (g) Ausencia de voz. (1. Glotis, 2. Cartílago 3. Tracto vocal, 4. Epiglotis) (Drugman et al., 2014).**

### 1.1.1.2 Modelos Matemáticos

Los métodos para modelos analíticos de la fuente glotal permiten entender algunos rasgos de la vibración glotal, pero no ayudan a identificar los que contribuyen a cualidades específicas de la voz relacionadas con emociones o algunos sonidos específicos. Muchos modelos de la fuente glotal han sido propuestos con la variación de varios niveles de complejidad como (Rosenberg, 1971, Fant.G, 1985, Fujisaki, 1986, Veldhuis.R, 1998, Titze, 1973, Marcelo de Oliveira Rosaa, 2003, Fariborz Alipour, 2011). Por lo anterior es que se emplean modelos matemáticos. Estos modelos fueron derivados de un análisis de medidas fisiológicas y alguno de ellos se muestran en los siguientes acápites.

El modelo Liljencrants-Fant (LF) fue presentado en 1985, en este se muestra (Figura.1.4) un ciclo de flujo glotal y su derivada de acuerdo a una voz modal. Se muestran las fases glotales más importantes (apertura, cierre y la fase de retorno). La fase de apertura se divide también en las fases apertura y cierre (Fant.G, 1985). Este ha sido usado en muchos artículos como (Ferreira, 2018, Aicha, 2018, Aníbal J. Ferreira, 2018).



**Figura.1.4. Formas de ondas típicas según el modelo LF. Flujo glotal (superior). Derivada del flujo glotal (inferior). También se indica el instante de cierre glotal (GCI) (Fant.G, 1985).**

Este modelo cuenta con cuatro parámetros que representan la primera derivada de la velocidad de volumen del pulso glotal, lo cual incorpora el efecto de radiación de los labios. Usa una combinación de funciones exponenciales y sinusoidales y es usado comúnmente para síntesis y análisis del habla (Fant.G, 1985).

El modelo Rosenberg fue presentado en 1971. En este se llevan a cabo dos experimentos. El primero consiste en usar diferentes formas de pulso (Figura.1.5) para la simular la excitación (Rosenberg, 1971). En total se utilizaron seis formas del pulso conformadas por triangular (Figura.1.5. A), polinomial (Figura.1.5. B), trigonométrica (Figura.1.5. C), trigonométrica (Figura.1.5. D), trigonométrica (Figura.1.5. E) y trapezoidal (Figura.1.5. F), todas ellas tenían fijo el tiempo de apertura y cierre (40 y 60% respectivamente).

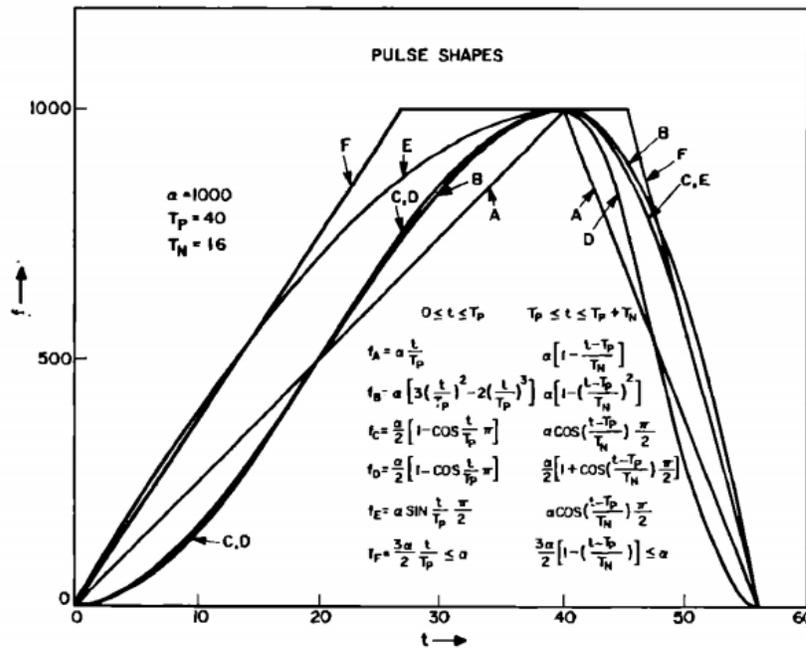


Figura.1.5. Formas de ondas utilizadas para estimulación (Rosenberg, 1971).

En el otro experimento solamente se consideró una forma (Figura.1.5.C) del pulso para simular la excitación, pero se tuvieron en cuenta cuatro rangos de apertura y cierre (9% hasta 60% para apertura y 4% hasta 50% para el cierre) teniendo un total de 16 combinaciones. Se llegó a la conclusión en el experimento 1 que las formas de ondas con mejor desempeño fueron las B y C y en el experimento 2 se encontró que existe una tolerancia para diferentes combinaciones de tiempos de apertura y cierre en donde estos sean grandes, pero con

pequeños tiempos de apertura o de cierre o con tiempos de apertura menores o aproximadamente iguales a los de cierre no se tuvieron buenos resultados.

El modelo Fujisaki-Ljungqvist (FL) fue presentado en 1986. Este modelo al igual que el LF representa la primera derivada de volumen del pulso glotal incorporando también el efecto de radiación de los labios. Cuenta con seis parámetros y funciones polinomiales, lo cual provee grandes detalles a la hora de modelar la forma del pulso glotal, pero es más difícil su uso en la práctica (Fujisaki, 1986).

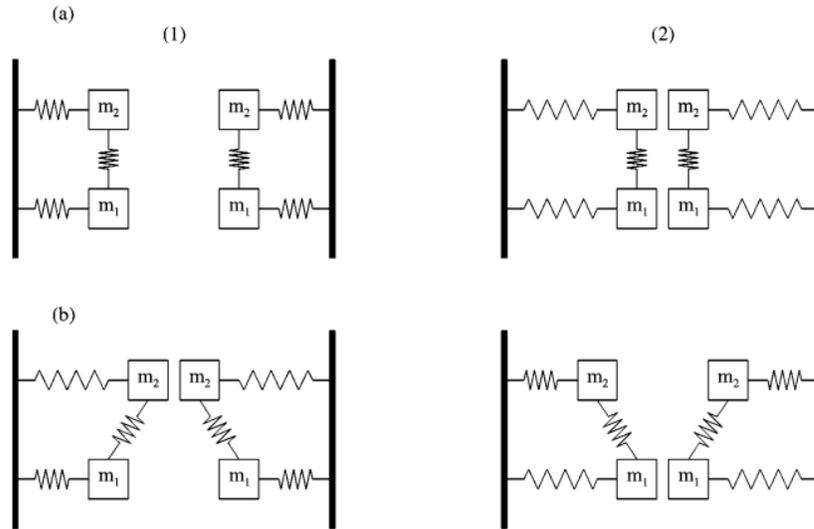
El modelo R++ fue presentado en 1998. Es una alternativa derivada del modelo de *Rosenberg*. Está descrito por los mismos parámetros  $T$  y  $R$  que el modelo LF, pero su ventaja es que es más eficiente computacionalmente, ya que computa directamente y de forma simple los parámetros antes mencionados. Se comparó con el modelo LF mediante un experimento acústico derivándose la conclusión que en situaciones prácticas es capaz de producir voces sintéticas perceptualmente equivalentes a las generadas por el modelo LF (Veldhuis, R., 1998).

En el año 2010 se introduce el modelo EE1 (Shue, 2010) motivado por la gran rapidez con que se obtienen las imágenes de cuerdas vocales, este tiene cuatro parámetros y usa una combinación de funciones exponenciales y sinusoidales similar al modelo LF, pero con la habilidad de ajustar las pendientes de las fases de apertura y cierre separadamente. En el año 2012 aparece el modelo EE2 (Chen, 2012), el cual es parecido al EE1 pero se redefinen los parámetros de velocidad de apertura y cierre para permitir una baja complejidad computacional, rápida generación de ondas y más exactitud a la hora de manipular la forma del pulso. Este último fue utilizado para la estimación automática del flujo glotal de señales de voz acústicas.

### **1.1.1.3 Modelos Físicos**

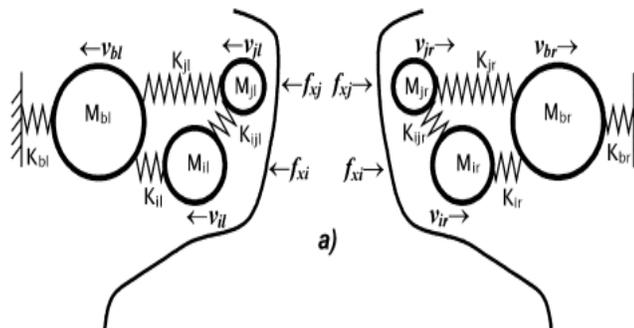
Respecto a los modelos físicos se utilizan distintos tipos de objetos para así modelar la fuente glotal. Estudios de resonancia de las cuerdas vocales proveen una investigación indirecta para entrenar modos básicos, indicando cuales modos básicos son más fáciles para entrenar basado en sus frecuencias naturales. El modelo de dos masas (Berry, 2001) puede ayudar a interpretar la estructura de resonancia de las cuerdas vocales. Está demostrado que muchos tipos de vibración de las cuerdas vocales pueden ser generados de bloques básicos conocidos

como modos básicos. El concepto de modos básicos puede ser más fácilmente explicado usando un simple modelo de bajo orden de las cuerdas vocales. El modelo físico más conocido de vibración de las cuerdas vocales es el de dos masas (Figura.1.6) de (Flanagan, 1975) el cual contiene dos modos básicos fundamentales.



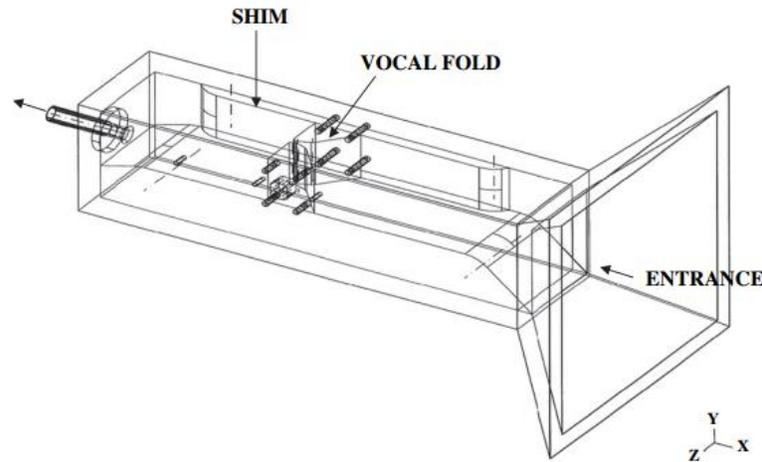
**Figura.1.6. Dos posibles modos básicos del modelo de dos masas, este incluye un modo básico con (a) las dos vibrando en fase y (b) las dos masas vibrando 180° fuera de fase. Las columnas 1 y 2 ilustran los modos básicos en extremos opuestos en el ciclo vibratorio. (Berry, 2001).**

En (Gómez-Vilda et al., 2007) se obtiene estimaciones de la correlación de la fuente glotal y su densidad espectral de potencia y singularidades, de estas estimaciones se deducen parámetros biomecánicos. En este método se usa el modelo k-masas (Figura.1.7) del sistema de cuerdas vocales ya que ayuda en la determinación y caracterización de los componentes principales de las cuerdas vocales en especial el efecto de la onda de mucosa. Al final se evalúan esos parámetros en la presencia de voces patológicas.



**Figura.1.7. Estructura esquemática de una sección del modelo de tres-masas (Gómez-Vilda et al., 2007).**

En (Fulcher et al., 2006) se utiliza un modelo estadístico físico de la laringe (M5) (Figura.1.8), el cual fue usado para obtener un conjunto de flujo de volumen como una función de la simetría geométrica glotal y presión transglotal. Se creó una ecuación generalizada para estimar la exactitud del volumen del flujo glotal, dando geometrías específicas glotales y presiones transglotales.

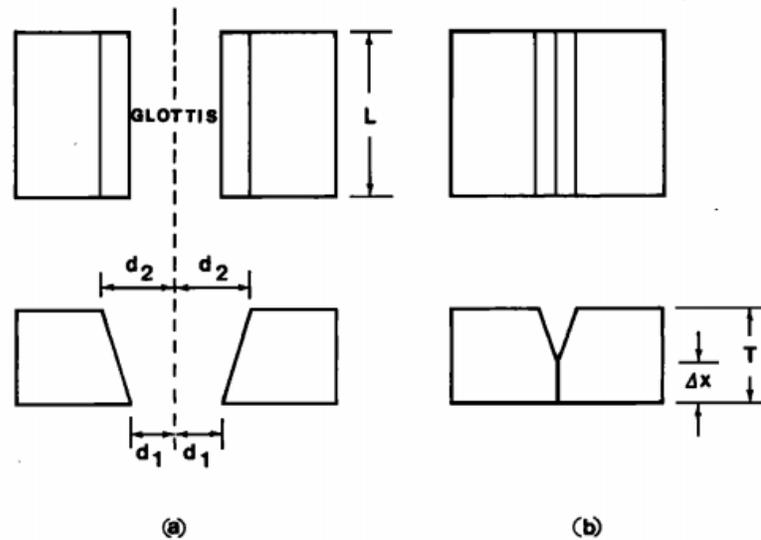


**Figura.1.8. Modelo esquemático M5 y cuerdas vocales (Fulcher et al., 2006).**

También se ha generado perturbaciones vocales como *jitter* usando modelos físicos (E. Cataldo 2016). En ese trabajo se construye un modelo estocástico para la creación de *jitter* utilizando un modelo mecánico de dos masas (Figura.1.6.), el cual ha sido usado para producir sonidos vocales. La rigidez del modelo físico fue considerada como un proceso estocástico, construyéndose con él una matriz estocástica. Esta matriz permitió la generación de *jitter* y su nivel fue controlado usando tres hiper-parámetros. Fue construida una función de densidad probabilística y evaluada para diferentes valores de los hiper-parámetros. El resultado final brindó excelentes resultados para la producción de *jitter* en las señales de voz.

Según (D.G. Childers, 1986) el modelo físico implementado en (Flanagan, 1975) tiene la ventaja de producir suficientes detalles para calcular la forma de onda de la voz, la distribución de la presión y de la velocidad del aire en el tracto vocal, observar la movilidad de las cuerdas vocales, el área glotal, entre otras. Una limitación de (Flanagan, 1975) es que el área glotal proyectada siempre es rectangular, por lo no es adecuado para el modelado real

de la vibración de las cuerdas vocales, por lo que se introdujo un modelo más realista con una masa elástica (Figura.1.9).



**Figura.1.9. Modelo elástico de una masa de las cuerdas vocales. a) Cuerdas vocales abiertas. b) Cuerdas vocales parcialmente cerradas (D.G. Childers, 1986).**

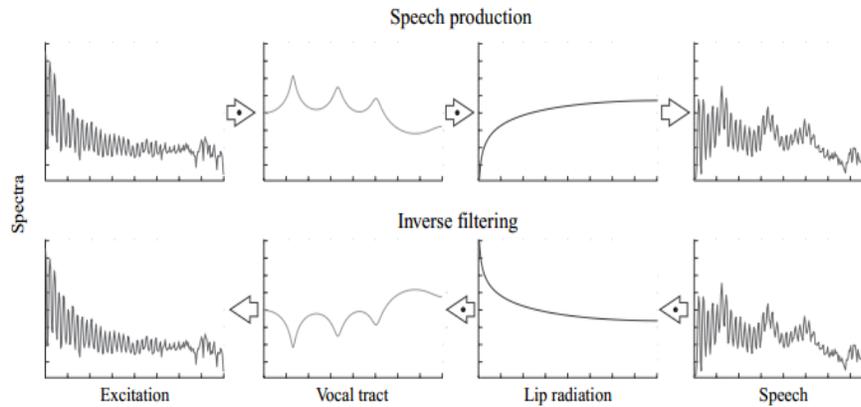
Se concluyó que este modelo no incluye todas las características de la laringe y que es mejor para voces masculinas en tono normal.

### 1.1.2 Modelos del Tracto Vocal (Filtro)

El filtro en el modelo Fuente-Filtro (G.Fant, 1960) es el tracto vocal, se han desarrollado numerosas técnicas para su conocimiento. Este modula la señal proveniente de la glotis para convertirla en sonidos.

El uso de un filtrado inverso es un método muy usado para estimar la fuente del sonido, la velocidad de onda del volumen glotal. Un modelo para el tracto vocal es computado primeramente en un filtrado inverso, estos métodos pueden ser divididos en dos variantes; la primera categoría está basada en la aplicación de una máscara (Alku and Laukkanen, 1998), la cual fue presentada por Rothenberg en 1973, la segunda categoría está basada en el procesamiento de la forma de onda de presión de la voz que ha sido grabada en el exterior de la boca. En (Airas, 2008) se crea un software donde se desarrolla un nuevo tipo de filtrado inverso de la voz y una parametrización. En este se utiliza la técnica de filtrado inverso

introducido en (Miller.RL, 1959) para remover los efectos del tracto vocal y los labios (Figura.1.10).



**Figura.1.10. Esquema de producción de la voz (superior). Técnica de Filtrado inverso (inferior) (Airas, 2008).**

Se ha explotado el dominio de la frecuencia para desarrollar métodos para determinar la función de transferencia del tracto vocal. En (Akande and Murphy, 2005) se explota la habilidad que tiene el espectro de tiempo corto para contener información sobre la fuente glotal y el tracto vocal. En el habla existe una apertura pequeña de baja frecuencia entre los picos de respuesta de frecuencia de la fuente glotal y el primer formante del tracto vocal.

En (Alku et al., 2009) se utiliza el análisis de covarianza de la fase-cerrada, el cual es otro algoritmo basado en el filtrado inverso que estima el tracto vocal mediante una predicción lineal durante la fase cerrada de la vibración vocal.

En (Huiqun Deng, 2006) se desarrolla un nuevo método para estimar el tracto vocal y la forma de onda glotal de vocales. En este se construyen dos ecuaciones (Ec.1.2 y Ec.1.3) para relacionar el tracto vocal con la señal de sonido y la forma de onda glotal con uno de los dos intervalos de cierre glotal.

$$M = \frac{2 \times L_{VT} \times F_s}{c} \quad (1.2)$$

$$H_{VTF}(z) = \frac{U_{lip}(z)}{(U_1^+(z) + U_1^-(z))} \quad (1.3)$$

En la Ec.1.2. se modela el tracto vocal usando un tubo cilíndrico de  $M$  secciones.  $L_{VT}$  es la longitud del tracto vocal,  $f_s$  es la frecuencia de muestreo y  $c$  es la velocidad del sonido. En la Ec.1.3 se define la función de transferencia del tracto vocal. Donde  $(U_1^+(z) + U_1^-(z))$  son la velocidad positiva y negativa del volumen y  $U_{lip}$  es la transformada de Fourier del volumen cuando los labios se abren. Después se restan las ecuaciones obtenidas donde se eliminan los componentes periódicos de la glotis de la estimación del tracto vocal, por lo que se obtiene un filtro vocal estimado. El promedio de muchas estimaciones de diferentes intervalos de cierre glotal conforma el estimado final, el cual es usado para obtener la forma de onda glotal por un filtrado inverso.

## 1.2 Perturbaciones Vocales

El *jitter*, *shimmer* y las perturbaciones compuestas son fenómenos que existen en todas las señales de voces reales. Estas son perturbaciones pequeñas, aleatorias en el tiempo y que varían en la amplitud y en el período. Si se examinan dos períodos consecutivos no van a ser iguales en términos de período, amplitud y complejidad.

El *jitter* es definido como una perturbación ciclo-ciclo en la frecuencia fundamental de las cuerdas vocales (Titze, 1995, Wendahl.W.R, 1966). Una de las muchas expresiones (Ec.1.4) para medir *jitter* se muestra a continuación:

$$\alpha = \frac{1}{N-1} \sum_{n=1}^{N-1} \frac{|T(n+1) - T(n)|}{0.5 \times (T(n+1) + T(n))} \times 100 \quad (1.4)$$

Donde  $T(n)$  es la longitud de  $n^{th}$  pulsos glotales entre los  $N$  pulsos analizados. La presencia de *jitter* ha sido importante en el análisis y síntesis de voces. En el análisis, pequeños valores de *jitter* son encontrados en voces normales, mientras valores altos son frecuentemente relacionados con voces patológicas. En síntesis, la introducción de pequeñas perturbaciones en la longitud del ciclo ha logrado que estas se acerquen más a las voces reales. La introducción de *jitter* en los sintetizadores de voz ha sido adicionalmente usado para explorar la percepción de aspectos acústicos de calidad vocal, un aspecto importante en la

evaluación de clasificaciones por escala. Valores de *jitter* menores del 1% son permisibles para voces normales. El *jitter* es afectado por la falta de control en la vibración de las cuerdas vocales.

El *shimmer* es una medida de variación ciclo a ciclo en la frecuencia fundamental donde lo que varía es la amplitud del pulso. La perturbación en la amplitud sirve como un índice de estabilidad vocal. El *jitter* y *shimmer* son comúnmente medidos en vocales sostenidas y su percepción sobre un cierto umbral está relacionado con voces patológicas como aspereza, jadeo, etc. Cuando están presentes distintos estilos del habla se ha reportado que ocurren también estos fenómenos en especial el shimmer. El *shimmer* excesivo está relacionado con la aspereza. Se considera normal una diferencia en la amplitud de 0.7dB o menos del 7% de la media de la amplitud. El *shimmer* es afectado por la reducción en la resistencia de la glotis y lesiones en las cuerdas vocales, lo anterior está relacionado con la presencia de ruido.

Para el estudio de todas estas perturbaciones se requiere primeramente la determinación del período fundamental en la señal o algoritmos de detección del período fundamental. Estos tipos de algoritmos se han desarrollado en el dominio de la frecuencia o en el tiempo (Asoke Kumar Datta, 2019).

### 1.3 Influencias, Enfoques y Contradicciones de los modelos de inducción de *jitter*

En este epígrafe se analizarán las diferencias, enfoques y contradicciones de varios modelos usados a la hora de sintetizar señales.

#### 1.3.1 Modelo *Constant-Warping*(CW)

En este modelo el pulso glotal se contrae o expande como un todo. La relación entre ambos pulsos se comporta como una línea recta con pendiente igual a  $T0^*/T0$ . Cualquier duración original de la fase  $T_x$  cambia en  $T_x^*$  de acuerdo a la Ec 1.5:

$$T_x^* = T_x \frac{T0^*}{T0} \quad (1.5)$$

En este modelo la versión *jittered* del pulso glotal es obtenida modificando las duraciones  $O_p$  y  $C_p$  en la misma proporción. La variante CW-GOI es frecuentemente implícita en la representación de  $g(t)$  en la literatura (Wendahl.W.R, 1966, J.M, 1999), en algunas

derivaciones teóricas (C. Manfredi, 2012) y en algunas implementaciones internas de sintetizadores (Dejonckere.H.P, 2012, C. Manfredi, 2012) de voces patológicas.

Este modelo trae consigo varios problemas como el relacionado con la definición de la cercanía del pulso. Para este modelo la longitud medida entre dos instantes adyacentes de apertura glotal difiere de la longitud medida entre dos instantes adyacentes de cierre glotal. Este problema es importante a la hora de evaluar el desempeño de algoritmos de detección del período fundamental (PDAs) cuando este modelo es asumido.

Un segundo problema para este modelo es que el uso de la Ec.1.1 en presencia de *jitter* produce una amplitud variable. Esto es debido a la modificación de la amplitud, la cual se contrae y expande cuando la derivada es aplicada. Pulsos cortos producen grandes amplitudes y viceversa. Este factor complica la estimación del *shimmer* glotal si esta opción es asumida (Ferrer et al., 2015).

### 1.3.2 Modelo *Piecewise-Warping*(PW)

En este modelo la pendiente de la línea referente a la fase de apertura de cualquier par de pulsos glotales es igual a uno. Este modelo usa un tren de impulsos como excitación a la función de transferencia del tracto vocal  $h(t)$ , con esta aproximación todas las fases diferentes de cero de pulsos glotales son iguales. Se ha usado con frecuencia en síntesis de *jittered* para evaluar el desempeño de PDAs o estimadores de HNR. Esta es la alternativa usada en el sintetizador de Klatt (Klatt, 1980b) y una implementada en el software Praat (Boersma and Van Heuven, 2001).

Una razón que favorece a esta opción es su simplicidad para sintetizar. Otra razón es la equivalencia de la longitud de los pulsos cuando es considerada la fase de apertura o cierre como puntos iniciales del ciclo. La inducción de *shimmer* por *jitter* también es menor en este caso. Por lo anterior el modelo CW puede ser considerado más cercano a la realidad que este (Ferrer et al., 2015).

## 1.4 Señal Electroglotográfica

La fotografía de alta velocidad ayuda a entender la diferencia en la vibración de las cuerdas vocales entre diferentes tipos de fonación, sin embargo, esta visualización no es posible

durante la producción natural de la voz, por lo que se usa la señal EGG por ser una técnica no invasiva y ser muy utilizada para analizar la voz (Mora, 2014).

#### 1.4.1 Principios Físicos

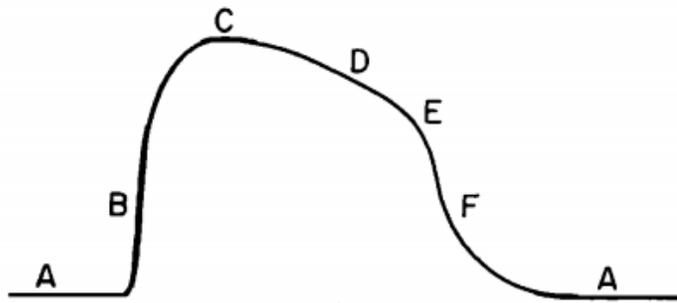
Para conseguir la señal EGG se deben colocar un par de electrodos (Figura.1.11) a ambos lados del cuello, en cada una de las láminas del cartílago tiroides, lo que permite, por medio de un filtro de paso alto, convertir la señal de análoga a digital y, a través de un software de compensación, obtener una onda electroglotográfica analizable.



Figura.1.11. Posición de los electrodos en el EGG (Mora, 2014).

Para comprender los fundamentos de este examen es necesario recordar la ley de Ohm: «la corriente eléctrica que circula entre 2 puntos pasa a través de un conductor y es proporcional al voltaje que la empuja. Una corriente constante disminuye cuando aumenta la resistencia, y viceversa. Así, la corriente es proporcional al voltaje e inversamente proporcional a la resistencia. Bajo esta lógica, la corriente que emite el electroglotógrafo por medio de uno de los 2 electrodos se transmite con facilidad a través de los pliegues vocales a una intensidad máxima de 10 mA, con una diferencia de potencial menor de 1 V, con lo que se evita la sensación de paso de corriente y no tiene efectos nocivos en el sujeto evaluado; dicha corriente es recibida por el otro electrodo, que lleva la señal a ser filtrada para generar una onda electroglotográfica. En el momento que los pliegues vocales se separan la corriente no se transmite, ya que el aire es un mal conductor eléctrico, lo que genera en la onda EGG un cambio en la amplitud debido al aumento de la impedancia o pérdida de la conductividad de la señal eléctrica emitida por los electrodos. Por el contrario, cuando se aproximan los pliegues vocales, la onda EGG muestra los cambios de impedancia eléctrica ante el estímulo

de corriente presentado. En otras palabras, cuando la corriente pasa de un electrodo, al contrario, atravesando de lado a lado los pliegues vocales en contacto, la onda obtenida llega a su nivel máximo de amplitud, debido al alto grado de conductividad que tiene la corriente eléctrica emitida. Mientras que al momento en que los pliegues vocales se separan, la corriente no pasa de un electrodo al otro y, por tanto, la onda generada tendría una amplitud mínima, quedando plana en la línea de base debido al aumento de impedancia para el paso de la corriente eléctrica. La señal EGG consta de varias fases como se muestran en la Figura 1.12.



**Figura.1.12. Fases de una señal EGG (Mora, 2014).**

El segmento A en la Figura 1.12 representa una línea base que refleja la máxima impedancia glotal asociada con la apertura glotal. El segmento B representa una pendiente creciente que representa un rápido decremento en la impedancia glotal y un progresivo contacto en las cuerdas vocales. C representa un pico lo que es equivalente a un mínimo en la impedancia glotal y un máximo en el contacto de las cuerdas vocales. D es la mitad descendente de la pendiente asociada con un incremento gradual en la impedancia glotal, ya que las cuerdas vocales se empiezan a separar. E es el ángulo de la EGG o la discontinuidad convexa que aparece el parte descendente de la onda. F es un segmento descendente después de la discontinuidad (Mora, 2014).

#### **1.4.2 Detección de instantes de interés**

La detección de instantes de interés (GCI- instante de cierre glotal, GOI- instante de apertura glotal) es muy importante a la hora de procesar la fuente glotal (Figura.1.13). En fonación normal  $g(t)$  es casi-periódica y por ende el tiempo entre estos instantes correspondiente a dos pulsos adyacentes es constante e igual al período de la señal ( $T_0$ ). En voces patológicas es

frecuente encontrar que  $T_0$  varíe más del 1% entre dos pulsos consecutivos, cuando esto pasa la equivalencia de usar GOI o GCI para medir el  $T_0$  no está garantizada. La forma de medirlos se explica en la Figura.1.2 una denotada  $T_{O-C}$ , lo cual denota un  $O_p$  y su posterior  $C_p$  (GOI) y otro denotado  $T_{C-O}$ , lo cual denota un  $C_p$  y su posterior  $O_p$  (Ferrer et al., 2015).

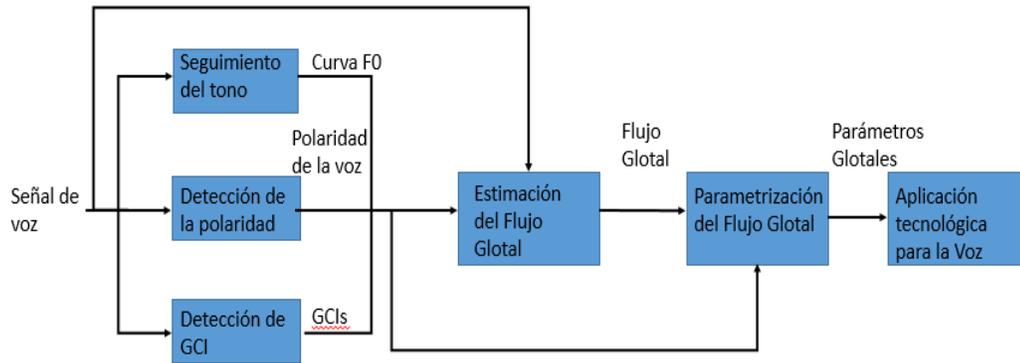


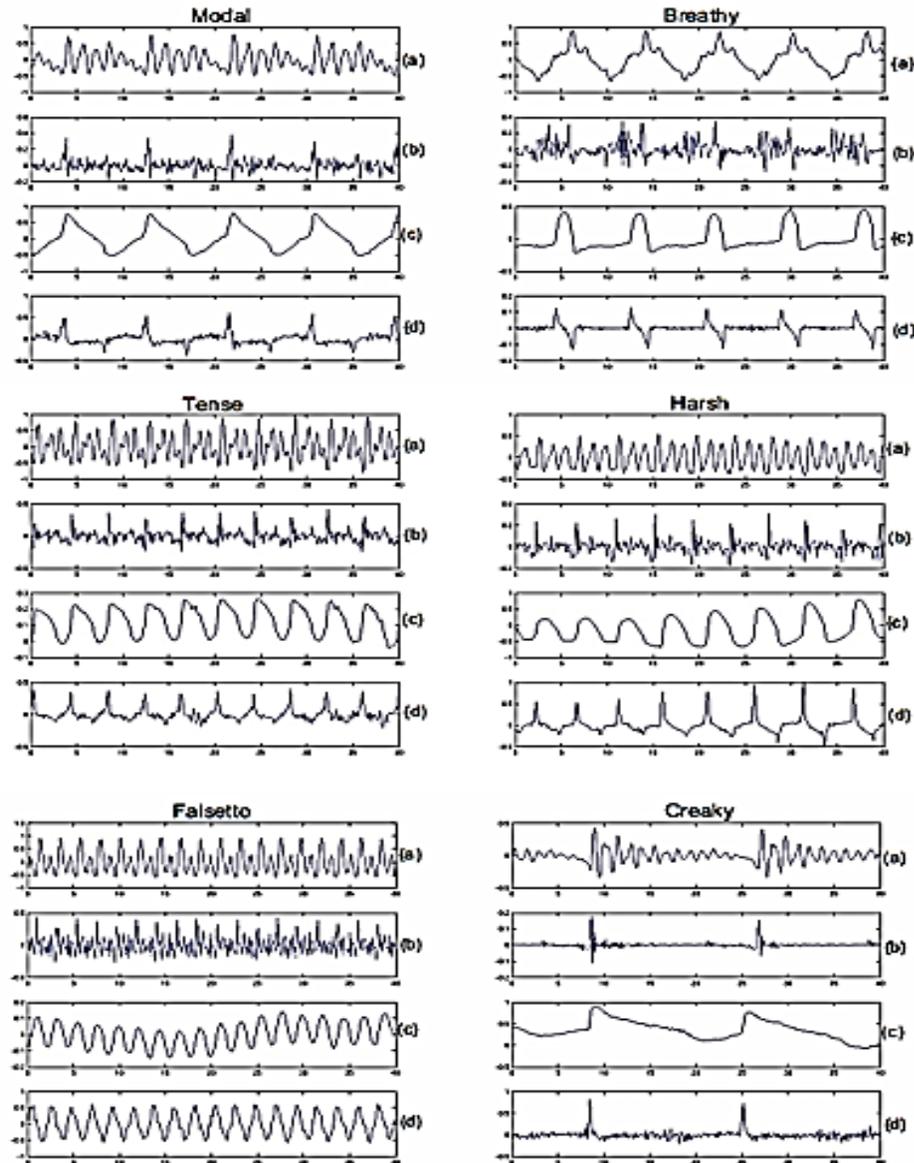
Fig.1.13. Sistema de procesamiento de la fuente glotal (Drugman et al., 2014).

#### 1.4.2.1 Detección de Instante de Cierre Glotal (GCI)

El instante de cierre glotal (Figura.1.2) es el paso de  $O_p$  a  $C_p$ . Las variaciones de la presión acústica causada por la vibración glotal que proviene del flujo de aire de los pulmones puede ser vista como la mayor excitación de la velocidad del sistema. Esta señal de excitación es filtrada por la respuesta del tracto vocal, generando la señal de voz. Si el componente de excitación es solamente un impulso, entonces la señal de voz corresponde a la respuesta del tracto vocal. La naturaleza del impulso como forma es perceptualmente importante y brinda una indicación de fuerza y sonoridad. En adición el comportamiento de la fase de retorno de la fuente glotal y la presencia de otras fuentes secundarias de excitación también contribuye a la calidad vocal del sonido (Drugman et al., 2014).

El conocimiento de GCIs es útil en muchas situaciones de análisis de voz como detecciones de regiones de actividad glotal, estimación de  $F_0$ , estimación de las frecuencias formantes, caracterización del sonido, análisis de señales de risa, extracción del tono de múltiples datos de voz, determinación del número de hablantes de señales mixtas, realce del habla en simples y múltiples casos, separación de múltiples hablantes, síntesis de voz, modificación de la prosodia, entre otras (Drugman et al., 2014). Existen muchos tipos de fonación (Figura.1.14)

que involucran diferencias en la vibración glotal, teniendo diferentes características en la cercanía del GCI.



**Figura.1.14. Formas de ondas. (a) Residual LP, (b) EGG, (c) dEGG, (d) señales para diferentes tipos de fonación: modal (superior izquierda), suspiro (superior derecha), tensa (medio izquierda), áspera (medio derecho), falseto (inferior izquierda) y voz chirriante (inferior derecha) (Drugman et al., 2014).**

La señal capturada por un EGG permite la extracción correcta de la localización de los GCIs y de la forma de la fuente de señal cercana a los GCIs, sin embargo, existen métodos para

extraer directamente GCIs de señales de voz y usar los detectados en las EGG para compararse. Estos métodos están agrupados en cinco categorías:

- 1- Métodos los cuales explotan la propiedad de la excitación de los GCI al impulso.
- 2- Métodos basados en propiedades de retraso de grupo (*group-delay*).
- 3- Métodos basados en el predictor lineal *all-pole*.
- 4- Métodos basados en la energía de tiempo corto de señales de voz.
- 5- Métodos basados en una combinación de muchas técnicas.

Método 1: Esta categoría pasa la señal de voz a través de un sistema que reduce el efecto de respuesta del tracto vocal, al mismo tiempo preserva la excitación al impulso propiedad en los GCIs. Uno de estos métodos es el resonador de frecuencia cero (ZFR), donde la señal de voz es pasada a través de una cascada de dos resonadores digitales ideales a 0 Hz, esto logra que los efectos de alta frecuencia de resonancia sean reducidos significativamente (Murphy.J.P, 2008, Drugman et al., 2014). La tendencia en la salida de ZFR es removida por la sustracción del promedio sobre una ventana del tamaño en el rango de uno a dos períodos de F0, este promedio brinda las localizaciones GCI precisamente en el instante de cruce por cero de negativo a positivo. En (Tuan, 1999) se utiliza la transformada de wavelet para la detección de GCI.

Método 2: Este grupo usa la derivada negativa de la fase de la transformada de Fourier de la señal. El promedio de cada segmento de la ventana de la señal corresponde al retraso al impulso en el segmento desde el centro del segmento. Entonces el promedio de la función *group-delay* computada en cada muestra tiene un valor de cero en el GCI, cuando este está localizado en el centro del segmento en análisis (Smits and Yegnanarayana, 1995). Se han refinado muchos algoritmos derivados de este como (Naylor, 2007) llamado DYPSA.

Método 3: Existen métodos que se basan en la detección de discontinuidades en la salida de un modelo de predicción lineal (LP) (Strube, 1974). Esta detección de GCI usando el modelo LP se realiza computando la envolvente de Hilbert, donde picos son obtenidos en la cercanía del GCI para señales limpias y esos picos coinciden con las discontinuidades encontradas en el diferencial EGG (dEGG) (Ananthapadmanabha and Yegnanarayana, 1979), existen refinamientos a estos algoritmos como (Kane and Gobl, 2013).

Uno de los errores del algoritmo LP es que los residuos a menudo contienen efectos de resonancia del sistema del tracto vocal, ya que el filtro inverso no cancela la resonancia

completamente. Este filtro reduce la SNR, lo cual resulta en una pérdida de resolución al impulso en la cercanía del GCI (Drugman et al., 2014).

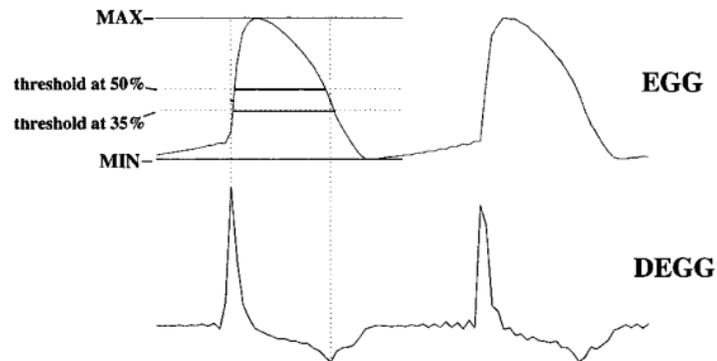
Método 4: Estos están basados en la energía de tiempo corto de la señal de voz o rasgos basados en la representación tiempo-frecuencia (Jankowski, 1995, Ma, 1994). Este método sufre de no ser muy exacto ya que requiere un bloque de procesamiento.

Método 5: En esta apartado se utilizan la combinación de varios métodos para la detección de GCI. El algoritmo YAGA usa un filtrado inverso adaptativo, análisis usando wavelet, *group-delay* y otros pasos (Thomas, 2012).

El algoritmo *SEDREAMS* es otro método que usa la media de la señal para determinar intervalos cortos donde GCIs es esperado y luego se asigna a cada intervalo una estimación más refinada para detectar los GCI usando residuos LP (Drugman, 2009). Este algoritmo fue posteriormente refinado para detectar GCI de diferentes tipos de fonación (Kane and Gobl, 2013).

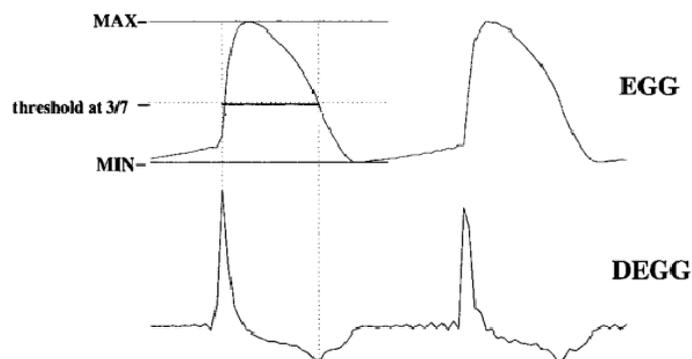
#### **1.4.2.2 Detección de Instante de Apertura Glotal (GOI)**

Aunque no es muy usual por parte de los investigadores por ser más difícil a la hora de detectar estos instantes. Algunos autores conectan el GOI al valor extremo de signo opuesto de la derivada del GCI. Sin embargo este punto no está bien definido como lo está el GCI debido a su transición es más suave que el GCI. Se han implementados métodos basados en umbral como el propuesto en (Rothenberg, 1988) donde se determina un nivel basado en un porcentaje de la amplitud entre el mínimo y el máximo de toda la señal glotal (50% para voces normales y 35% para voces relajadas). Los puntos que cruzan en la Figura 1.15 desde la señal EGG a su derivada son aproximadamente los instantes de cierre y apertura glotal. Este método es muy conveniente para propósitos médicos y puede ser aplicado a señales ruidosas y débiles, pero su resultado es impreciso y puede ser inexacto comparado con lo medido en una señal de flujo glotal o en una señal de área glotal (Rothenberg, 1988).



**Figura.1.15. Método basado en umbral implementado en (Rothenberg, 1988).**

Una alternativa reportada para la localización del GOI que brinda buenos resultados es la localización cuando la onda desciende  $3/7$  (Howard, 1995) de toda la extensión como se muestra en la Figura. 1.16. En esa investigación se usa el pico de cierre de la señal DEGG para detectar el instante de cierre glotal y el umbral antes mencionado en la señal EGG para detectar el instante de apertura glotal. Este método brinda resultados más precisos, por la detección más cercana del instante de cierre glotal.



**Figura.1.16. Método basado en el umbral de  $3/7$  de la señal (Howard, 1995).**

El algoritmo YAGA usa un filtrado inverso adaptativo, análisis usando wavelet, *group-delay* otros pasos (Thomas, 2012), aunque se usó para GCI, también fue usado para GOI. En (Bouزيد, 2004) detectan GOI usando la transformada de wavelet, pero no tuvo buenos resultados. Para la detección de GOI también se usó el algoritmo SEDREAMS (Drugman, 2009), el cual demuestra más robustez al ruido que previas investigaciones. En este artículo

se observa que la exactitud con que se detecta el GOI depende las características de la fuente glotal.

En (Sara Anastaplo, 1988) se utilizaron dos técnicas para la obtención de la apertura glotal, estas fueron la señal EGG y la videostroboscopia. Las medidas en la investigación realizada fueron pronunciando la vocal |i| por más de 5s. La técnica de videostroboscopia sirvió para marcar tangentes en las partes más importantes de la señal EGG (Figura.1.17) permitiendo la precisa identificación de la discontinuidad en la EGG.



**Figura. 1.17. Tangentes marcadas en la señal EGG.**

### **1.5 Conclusiones Parciales**

En este capítulo se abordaron los diferentes modelos matemáticos, físicos y fisiológicos de la fuente glotal, así como los modelos del tracto vocal. Se describieron los principios de la señal EGG y como es muy usada por ser no invasiva y describir a la fuente glotal. Se analizaron las fases de la señal EGG, se determinó que el instante de apertura glotal es muy difícil de encontrar debido a sus características. Ningún modelo existente permite modelar las longitudes de las fases del pulso con respecto a la duración total de este. No se ha evaluado la correspondencia de señales reales con algún modelo de relación.

## CAPÍTULO 2. Correlaciones asumidas entre las fases y la longitud del pulso glotal

En este capítulo se abordará sobre la base de datos de señales utilizadas, la selección de segmentos en el primer epígrafe. En el segundo epígrafe se profundizará en la detección de GCI y GOI. Seguidamente se hallarán las medidas de relación entre el pulso glotal y las fases. Se expondrán los modelos de relación considerados y se arribarán a conclusiones parciales.

### 2.1 Señales EGG empleadas

En este trabajo se utilizaron señales EGG debido a que existe más disponibilidad de estas respecto a videoendoscopía o videocamografía. La duración de las señales en las últimas técnicas es usualmente muy estricta. La señal EGG puede ser considerada próxima a el flujo glotal ( $g(t)$ ), los puntos relevantes en el pulso glotal GOIs y GCIs son rutinariamente detectados en esta. Todos los métodos y experimentos están implementados en Matlab 2017.

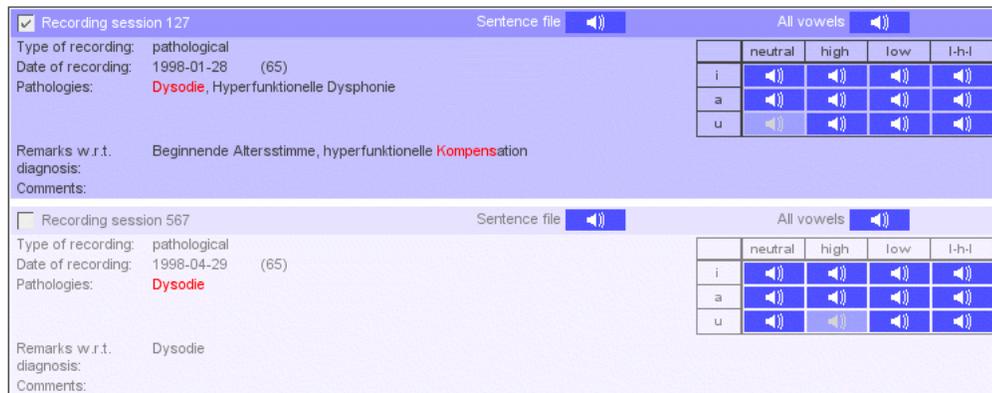
#### 2.1.1 Base de datos de voz *Saarbrücken*

La base de datos de señales usada en este estudio es extraída de la base de datos *Saarbrücken* (M. Pützer ), la cual está conformada por grabaciones de voces patológicas y normales de más de 2000 personas a una  $fs=50kHz$ . Una sesión de grabación contiene:

- Grabaciones de las vocales i, a, u producida a frecuencias normales, bajas y altas.
- Grabaciones de las vocales i, a, u con frecuencias que suben y bajan.
- Grabaciones de oraciones como “buenos días”, “como estas” en el idioma alemán.

Para lo anterior la señal de voz y la EGG están guardadas en archivos diferentes. En dependencia de la calidad de la grabación, pueden estar algunas o todas las vocales u oraciones antes expuestas. Esta base de datos tiene una interfaz web (Figura.2.1.) que le brinda acceso a esta. Esta interfaz contiene otras páginas que pueden ser usadas para:

- Selección de criterios para la base de datos.
- Reproducir directamente grabaciones.
- Escoger cuales archivos de las grabaciones se quieren importar.

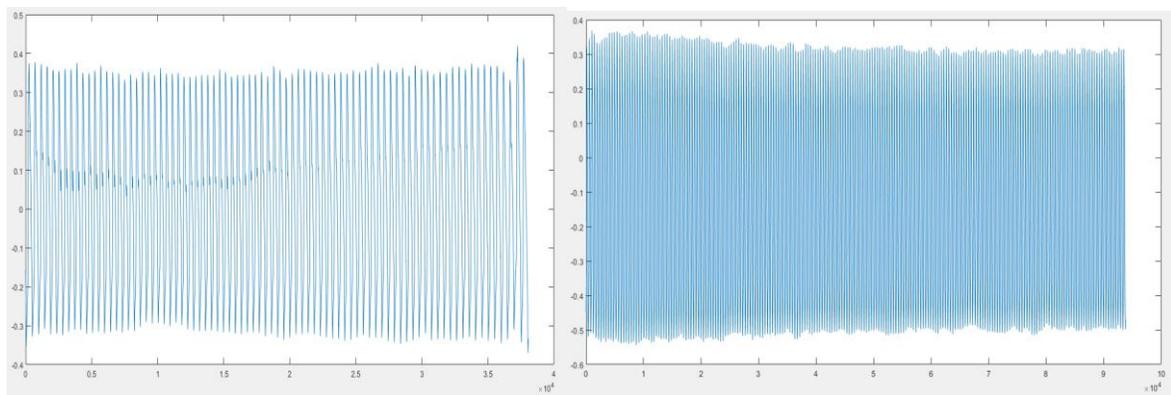


**Figura.2.1. Interfaz web de una sesión de la base de datos Saarbrücken (M. Pützer) .**

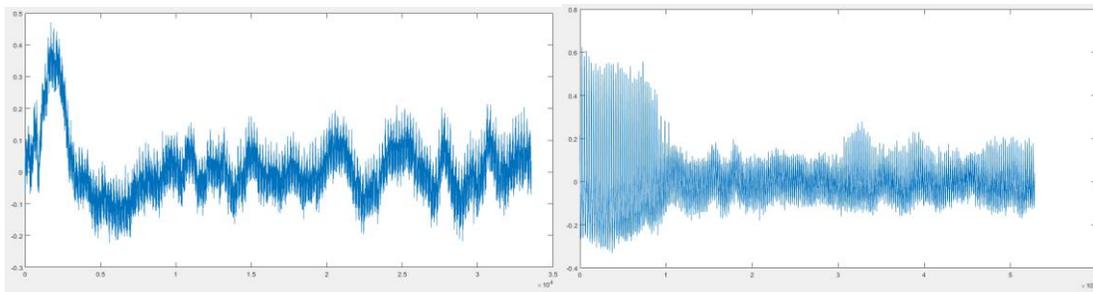
En este trabajo se escogió trabajar con grabaciones de señales EGG de la vocal /a/ en tono normal.

### 2.1.2 Selección de señales EGG

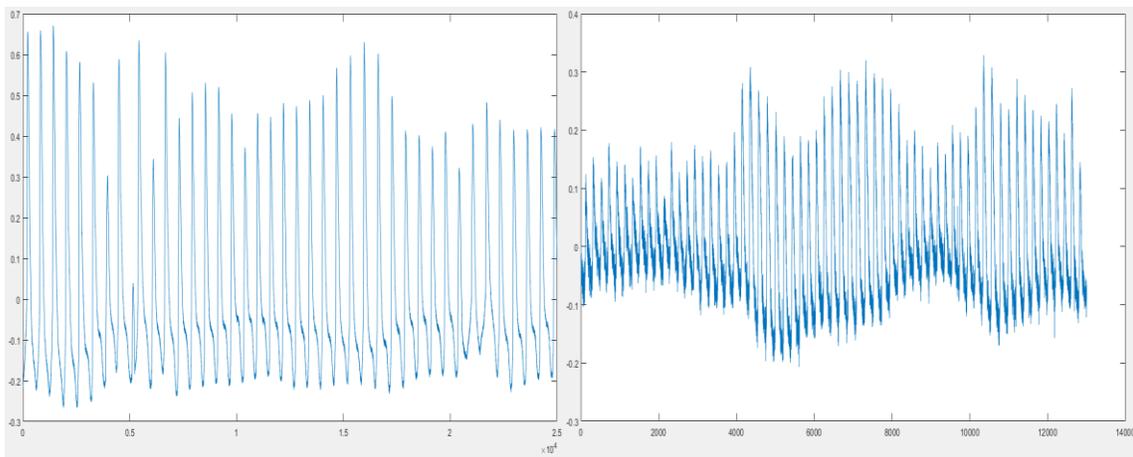
En esta investigación se seleccionaron 114 señales basándonos en dos aspectos: variabilidad e irregularidad. Si las señales no tienen variabilidad (Figura.2.2) o son irregulares (Figura.2.3) no se puede llegar a correlacionar las fases con la duración del pulso. Las señales escogidas (Figura.2.4) no tenían los problemas antes expuestos.



**Figura.2.2. Señales regulares. Señal 1426 (izquierda). Señal 1897 (derecha).**



**Figura.2.3. Señales irregulares. Señal EGG 1450 (izquierda). Señal EGG 108 (derecha).**



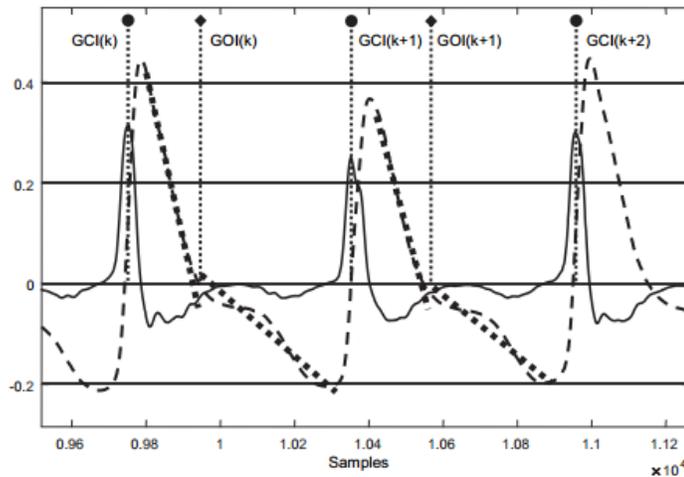
**Figura.2.4 Señales escogidas. Segmento de la señal EGG 110 (izquierda). Segmento de la señal EGG 715 (derecha).**

### 2.1.3 Selección de segmentos

Primeramente, se realizó una inspección visual en todos los casos patológicos para buscar segmentos que tuvieran un pulso evidente proporcionando variabilidad en estos con el fin de encontrar correspondencia en el análisis correlacional, seguidamente se buscó un patrón de pulsos bien definido permitiendo que el marcado automático se pudiera realizar. Después de esta inspección se seleccionaron 114 señales, de estos 120 segmentos EGG fueron extraídos. Se detectaron 5763 pulsos glotales, existiendo una variación en los segmentos que iba de un mínimo de 8 a un máximo de 302, para una media de 48. Los segmentos excluidos de cada señal fueron porque eran muy regulares para proveer una variabilidad útil o muy irregulares para el desempeño de las medidas requeridas.

## 2.2 Métodos usados para la detección de GCI y GOI

En las señales EGG como en  $g(t)$ , la ubicación de los GCI es donde la derivada alcanza su máximo valor. En las señales EGG ( $L_x$ ) usadas en este trabajo el pico en la derivada correspondiente al GCI en un pulso es el máximo positivo (Figura.2.5.), con un comportamiento opuesto respecto a  $g(t)$  en ambos tipos de señales EGG la parte más plana ocurre antes del GCI (correspondiendo a la fase Op) mientras la parte inferior de la fase Cp (después del GCI) presenta grandes bajadas y subidas.

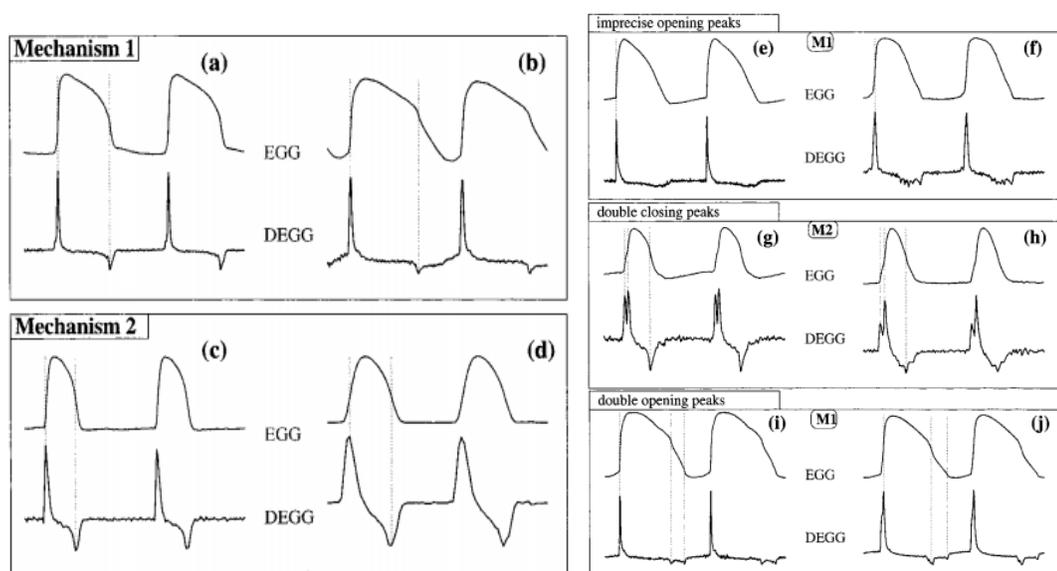


**Figura.2.5.** Señal  $L_x$  (líneas discontinuas), su derivada (líneas solidas) extraídas de un segmento de la señal 110 (eje vertical con unidades arbitrarias). GCIs es mostrado en círculos y GOIs en diamantes.

El método usado para la detección de GCI (Figura.2.5) fue pico a pico (Henrich.N, 2004). En (Henrich.N, 2004) primeramente se analizan varios casos en la derivada de la EGG, donde pueden ocurrir que aparezca picos bien definidos o no bien definidos en la fase de apertura o cierre (Figura.2.6.). Donde aparecen picos bien definidos, la fase de cierre relativa a la distancia entre dos líneas sucesivas verticales se puede tomar como la frecuencia fundamental. En donde aparecen picos no tan bien definidos se puede dividir en tres categorías: picos imprecisos en la fase apertura, doble picos en la fase de cierre y doble picos en la fase de apertura. En la Figura.2.6 en el caso e) no se detecta ningún pico en la apertura, en el caso f) se detectan numerosos picos y ninguno de ellos sobresale claramente. En el caso g) y h) aparecen dos picos en la fase de cierre. El fenómeno anterior también puede ser encontrado en la fase de apertura como se observa en los casos i) y j). Como se observa en los casos que existen doble picos en la fase de apertura respecto a los de la fase de cierre,

estos son menos fuertes y menos cercanos uno del otro. Se puede observar que los picos en la fase de apertura o cierre rara vez son triples o cuádruples.

Se especula que la aparición de dobles picos está relacionada con diferencias en el espesor de las cuerdas vocales en las fases de apertura o cierre o de un intervalo de cierre o apertura en dos partes diferentes de la glotis. Como conclusión de esta parte se llegó a que cuando los picos observados en la derivada de la señal EGG son simples, estos están relacionados con los instantes de cierre o apertura glotal. En este caso se pueden usar para medir la frecuencia fundamental.



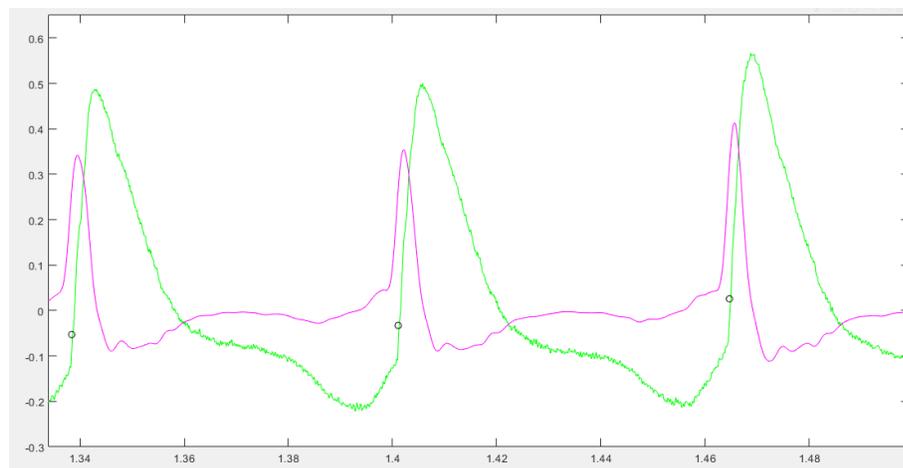
**Figura.2.6. Señal EGG (parte superior a, b, c, d, e, f, g, h, i, j), dEGG (parte inferior de la derivada de la señal EGG a, b, c, d, e, f, g, h, i, j) (Henrich.N, 2004).**

Este método llamado DECOM brinda buenos resultados cuando la fase de apertura y cierre están bien definidas. Cuando la fase de apertura no está bien definida, este ya no puede ser aplicado. Su principio se basa en la medición de la frecuencia fundamental y el cociente de apertura usando la correlación para estimar la distancia entre dos picos consecutivos en la fase de cierre y la distancia entre un pico en la fase de apertura y su consecutivo en la fase de cierre (Henrich.N, 2004).

Este fue aplicado a la versión suavizada de la señal EGG. Los picos fueron obtenidos cambiando tres parámetros de la función de Matlab *FindPeaks* en la derivada (Figura.2.7) de la señal EGG suavizada. Estos parámetros fueron:

- '*MinPeakDistance*': A  $1/3$  de un  $T_0$  estimado (obtenido como el inverso del primer armónico en un análisis FFT de toda la señal EGG).
- '*MinPeakProminence*': A  $(m_x - m_n) / 6$  (siendo  $m_x$  el máximo global de la señal EGG y  $m_n$  el mínimo global).
- '*MinPeakHeight*': A  $m_x / 12$ .

El suavizado fue desempeñado en un proceso de filtrado de dos pasos; primeramente, se aplica un filtro de mediana para remover posibles ruidos y segundo se aplica un promedio móvil con una ventana de *Hanning*. Ambos filtros operan sobre ventanas de 41 muestras (0.8ms), lo que significa que el promediado móvil representa un filtro paso-bajo con frecuencia de corte aproximadamente 850 Hz.

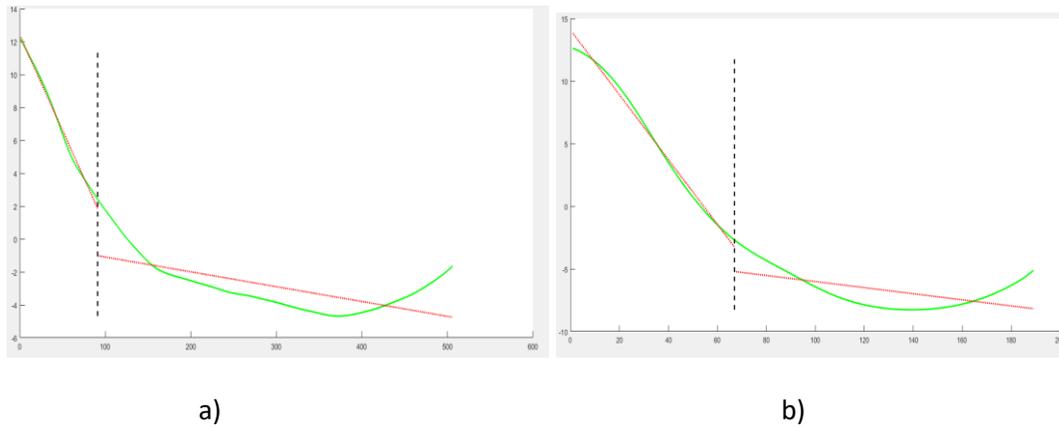


**Figura.2.7.** GCIs encontrados según el método (Henrich.N, 2004). Señal EGG (verde). Derivada de la señal suavizada (violeta). GCIs círculos negros.

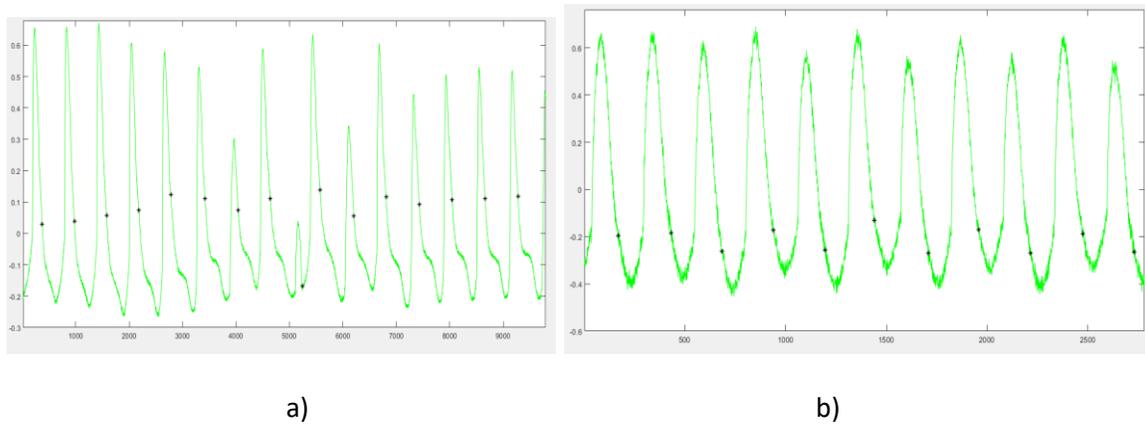
La detección de los GOIs representa un problema debido a que sus puntos no están bien definidos como lo están los correspondientes al GCI. La excursión de caída en donde se encuentra el GOI es significativamente menor que la excursión de subida donde el GCI es detectado, por esta razón es frecuente encontrar múltiples mínimos locales en la sección de caída (primera y tercera ocurrencia en la Figura.2.5.) y la posición en el tiempo del pico mayor puede variar con la sección de caída correspondiente.

En esta investigación se usó un método de optimización ajustando dos secciones de la onda EGG. Las secciones modeladas (Figura.2.8) son la excursión de caída dentro del  $C_p$  y la parte más plana que ocurre antes del próximo GCI. Cada sección es modelada como una línea recta, la que es obtenida de la asignación óptima de la onda EGG. Usando un valor de margen auxiliar de  $m=0.05(GCI_{(k+1)}-GCI_{(k)})$  las líneas son definidas en el intervalo  $[t_{peak} + m, t_{mid}]$  y

$[t_{mid}, GCI_{(k+1)} - m]$  respectivamente, siendo  $t_{peak}$  el máximo instante de la señal EGG entre  $GCI_{(k)}$  y  $GCI_{(k+1)}$ . El  $GOI(k)$  (Figura.2.9) es el valor de  $t_{mid}$  evaluado en el intervalo  $[t_{peak} + 3m, GCI_{(k+1)} - 3m]$  para el cual el error cuadrado total de optimización (considerando ambas líneas) es minimizado.



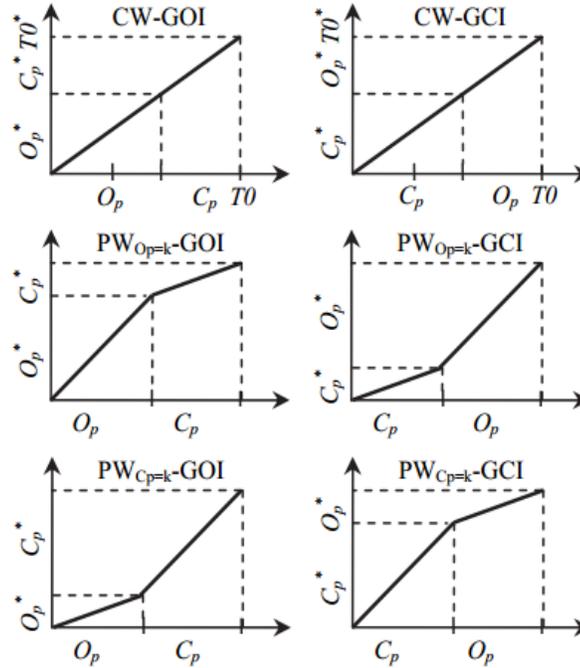
**Figura.2.8. Ajuste de un segmento de la señal EGG. a) Segmento 1 de la señal 1. b) Segmento 1 de la señal 112. Segmento de las señales en verde. Ajuste del segmento en color rojo. Recta vertical es el punto de discontinuidad.**



**Figura. 2.9. GOIs encontrados en la señal EGG. a) GOIs encontrados en la señal EGG 1. b) GOIs encontrados en la señal EGG 112.**

### 2.3 Modelos de relación considerados

Para evaluar la relación que existe entre la longitud del pulso glotal y sus fases  $Op$  y  $Cp$  se utilizaron tres modelos (CW, PW  $Op=k$  y PW  $Cp=k$ ) (Figura.2.7.). Las versiones de la parte izquierda de la Figura.2.10 corresponde a los GOIs de los tres modelos y la parte derecha a los GCIs.



**Figura.2.10. Modelos para el mapeo de la variación (jitter) de la longitud del pulso en las fases  $O_p$  y  $C_p$ . Modelo CW (superior), Modelo PW con  $O_p$  constante (medio), PW con  $C_p$  constante (inferior). Versiones basadas en GOI (izquierda) y GCI (derecha).**

El modelo *Constant-Warping* se usa para modificar la duración de las fases de pulsos individuales según el cambio en la longitud total del pulso. En este trabajo se usó combinado con las fases GCI y GOI. Otro modelo usado es *Piecewise-Warping* con  $O_p$  constante y por último el modelo *Piecewise-Warping* con  $C_p$  constante. Este modelo es una alternativa al PW, pero en este caso se asume constante a la fase glotal  $C_p$ . Este no ha sido descrito, representado o usado en ningún experimento hasta la fecha.

## 2.4 Medidas de relación entre pulso glotal y sus fases

En este epígrafe se explicarán las medidas y correlaciones encontradas entre el pulso glotal y sus fases.

### 2.4.1 Secuencias extraídas

De las secuencias  $GCI(k)$  y  $GOI(k)$  obtenidas y explicadas en el epígrafe 2.2 se derivaron cuatro secuencias para los 120 segmentos de las señales EGG.

- $C_{p(k)}$ : Tiempo decursado entre  $GCI(k)$  y  $GOI(k)$ .

- $O_{p(k)}$ : Tiempo decursado entre  $GOI_{(k)}$  y  $GCI_{(k+1)}$ .
- $T_{C0(k)}$ : Tiempo decursado entre  $GCI_{(k)}$  y  $GCI_{(k+1)}$ . Contiene a  $C_{p(k)}$  y su posterior  $O_{p(k)}$ .
- $T_{0C(k)}$ : Tiempo decursado entre  $GOI_{(k)}$  y  $GOI_{(k+1)}$ . Contiene  $O_{p(k)}$  y su posterior  $C_{p(k+1)}$ .

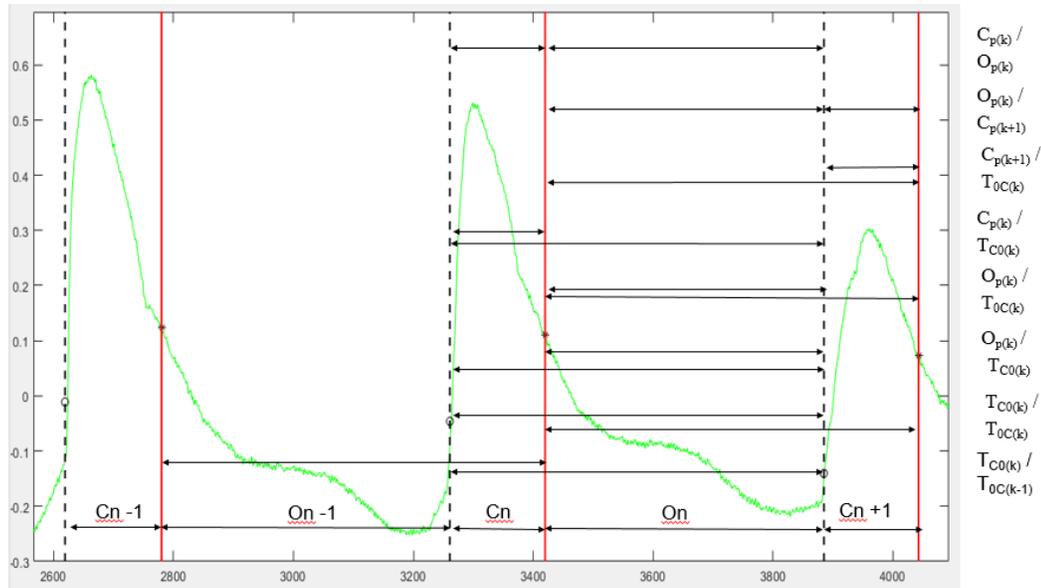
### 2.4.2 Correlaciones extraídas

Un total de 8 posibles relaciones fueron escogidos entre las secuencias antes explicadas, su significado y valores hipotéticos para cada modelo son mostradas en la Tabla.2.1. En la parte superior izquierda se encuentra la correlación  $TC0(k)$  vs  $T0C(k-1)$  la cual está fuertemente relacionada solamente si la duración está compartida,  $C_p$ , es la fuente principal de variabilidad, esto es verdadero solamente para el modelo PW  $Op=k$  y aparece marcado con una X en esta tabla. La segunda columna es similar, una correlación entre las longitudes del pulso que solo comparten  $Op$ , y debería ser grande para el modelo PW  $Cp=k$ . La tercera columna es para la correlación entre  $Op$  y la longitud del pulso definido como GCIs. Esta puede ser grande para CW-GCI, donde  $Op$  comparte la variabilidad con su  $Cp$  previo, o para PW  $Cp=k$  donde solamente varía la duración. La cuarta columna correlaciona  $Op$  con la longitud del pulso definido como GOIs, por razones equivalentes a lo explicado en la tercera comuna, se encuentra marcado el modelo CW-GOI y PW  $Cp=k$ . La quinta y sexta columnas repiten lo que sucede en la tercera y cuarta, pero se sustituye  $Op$  por  $Cp$ . Las dos últimas columnas exploran la correlación entre  $Op$  y  $Cp$  y se encuentra marcada solamente para el modelo CW. En la séptima columna  $Op$  precede a  $Cp$  y está marcada en el modelo del pulso glotal GOI, sucediendo lo contrario en la octava columna.

Como ha sido previamente mencionado no existen diferencias significativas en el modelo PW respecto a las cercanías del pulso (GCI y GOI) por lo que se han unido en las filas de la Tabla 2.1 correspondiente al modelo PW.

**Tabla 2.1. Matriz de correlación esperadas.**

Modelo	$T_{C0(k)} /$ $T_{0C(k-1)}$	$T_{C0(k)} /$ $T_{0C(k)}$	$O_{p(k)} /$ $T_{C0(k)}$	$O_{p(k)} /$ $T_{0C(k)}$	$C_{p(k)} /$ $T_{C0(k)}$	$C_{p(k+1)} /$ $T_{0C(k)}$	$O_{p(k)} /$ $C_{p(k+1)}$	$C_{p(k)} /$ $O_{p(k)}$
CW-GOI				X		X	X	
CW-GCI			X		X			X
PW $Op=k$	X				X	X		
PW $Cp=k$		X	X	X				



**Figura.2.11. Correlaciones supuestas.**

En la Figura 2.11 se muestran las correlaciones supuestas en una señal de muestra.

## 2.5 Conclusiones parciales

En este capítulo se describió a la base de datos utilizada para el análisis desarrollado en el capítulo 3. Se implementó un nuevo método para la obtención del GOIs y para la del GCI se usó uno que estaba en la bibliografía. Se explicaron los modelos de relación y se obtuvieron las correlaciones y secuencias entre estos y las fases GOI y GCI.

## CAPÍTULO 3. Evaluación de las relaciones encontradas entre las fases y la longitud del pulso

En este capítulo se analizará y evaluará la relación que existe entre las fases del pulso glotal y las longitudes los puntos iniciales y finales de la onda EGG.

### 3.1 Análisis matemático de las correlaciones

En este epígrafe se describirán las ecuaciones (Ec.3.1-3.4) que relacionan la longitud del pulso glotal con las fases.

$$CW - GOI = \frac{\sum Corr1 + Corr2 + Corr3}{3} \quad (3.1)$$

Donde:

$$Corr1 = (O_{p(k)}, T_{0c(k)}) , Corr2 = (C_{p(k+1)}, T_{0c(k)}) , Corr3 = (O_{p(k)}, C_{p(k+1)})$$

$$CW - GCI = \frac{\sum Corr1 + Corr2 + Corr3}{3} \quad (3.2)$$

Donde:

$$Corr1 = (O_{p(k)}, T_{c0(k)}) , Corr2 = (C_{p(k)}, T_{c0(k)}) , Corr3 = (C_{p(k)}, O_{p(k)})$$

$$PW_{Op=k} = \frac{\sum Corr1 + Corr2 + Corr3}{3} \quad (3.3)$$

Donde:

$$Corr1 = (T_{c0(k)}, T_{0c(k-1)}) , Corr2 = (C_{p(k)}, T_{c0(k)}) , Corr3 = (C_{p(k+1)}, T_{0c(k)})$$

$$PW_{Cp=k} = \frac{\sum Corr1 + Corr2 + Corr3}{3} \tag{3.4}$$

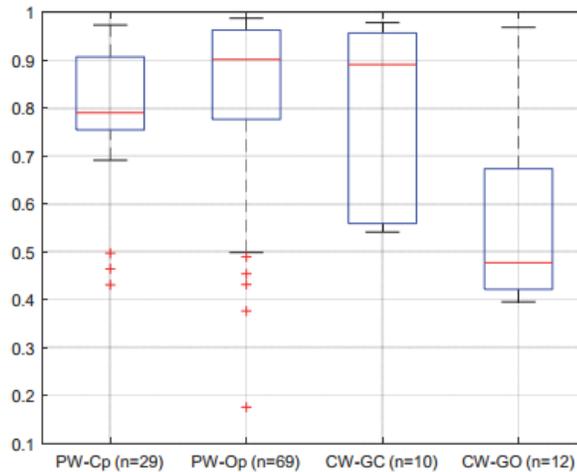
Donde:

$$Corr1 = (T_{C0(k)}, T_{0C(k)}) , Corr2 = (O_{p(k)}, T_{C0(k)}) , Corr3 = (O_{p(k)}, T_{C0(k)})$$

Todos los valores de las correlaciones que dieron por debajo de 0.5 fueron reemplazados por cero.

### 3.2 Fuerza de las perturbaciones

Los ocho valores de correlación descritos en la Tabla 2.1 fueron obtenidos para los 120 segmentos de las señales EGG. Se conformó una matriz de correlación de 120x8 donde se mantuvieron los valores positivos de moderado, fuerte y muy fuerte según (Ferrer et al., 2018). Consecuentemente todos los valores de *r Pearson* por debajo de 0.5 fueron reemplazados por cero como se abordó en el epígrafe anterior. La correspondencia de un segmento (fila) dado con un modelo fue determinada por la mayor suma de las correlaciones en la posición marcada por el modelo en la Tabla 2.1. Los resultados son mostrados en la Figura.3.1.



**Figura.3.1. Ploteo en forma de caja de la media de tres correlaciones significativas por modelo. El número de ocurrencia de cada modelo se encuentra entre paréntesis.**

### 3.3 Relaciones encontradas

Los resultados en la Figura.3.1 muestran la prevalencia del modelo PW donde toda la variabilidad es asumida por Cp con 69 ocurrencias (57%). Este modelo también muestra el valor medio más alto en los segmentos medios, sobre 0.9, sugiriendo que el efecto es realmente fuerte en esos segmentos donde aparece.

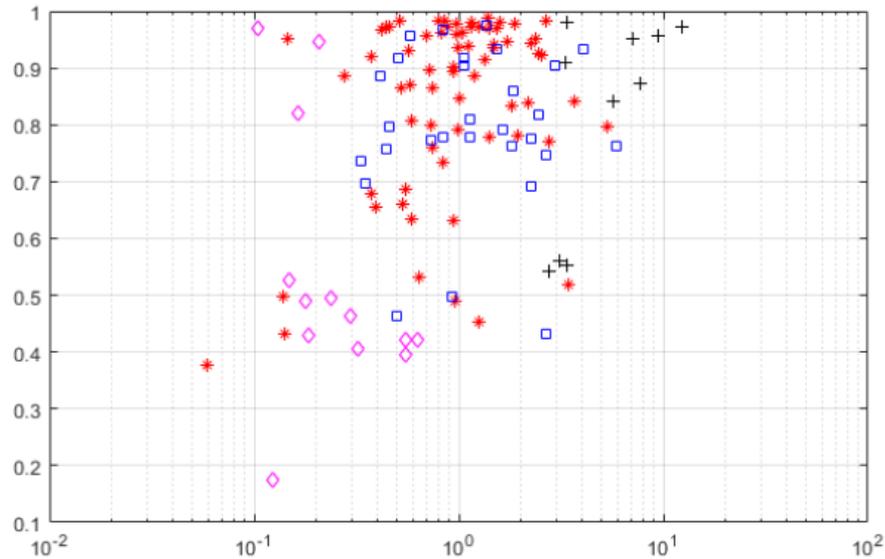
Aunque el modelo PW Op=k es el más frecuente y fuerte, la frecuencia de ocurrencia de los otros modelos no puede ser borrada. Menos de la mitad de frecuente y definitivamente débil el modelo PW Cp=k aparece (~25%). El modelo CW-GCI es el menos frecuente, pero cuando aparece es tan fuerte como PW Op=k. El modelo CW-GOI puede ser considerado como afectado por estos resultados, un descubrimiento controversial considerando la posición bien establecida de este modelo en la literatura y los sintetizadores.

### 3.4 Relaciones encontradas basadas en la cantidad de jitter

Otro criterio por el cual se puede estimar a que modelo de los considerados pertenece una señal real es el de la medición de jitter considerando una frontera o la otra (GCI o GOI). Cualquier modelo de los *PW-Warping* debe producir valores de jitter semejantes para ambas fronteras. Por otra parte, los modelos CW-Warping deben producir valores de jitter para sólo una de las fronteras. Por este motivo se decidió emplear la relación:

$$Relación = \frac{Jitter\_GCI}{Jitter\_GOI} \quad (3.5)$$

En la Figura.3.2 se observa como lo anterior expuesto que verificado. En los modelos PW tanto Cp=k como Op=k están alrededor de uno por aparecer en las fases GOI y GCI valores de jitter muy parecidos. En los modelos CW-GOI los valores están por debajo de uno por prevalecer el jitter en la fase GOI. En los modelos CW-GCI los valores están por encima de uno por prevalecer el jitter en la fase GCI.



**Figura.3.2. Ploteo de ocurrencia de los modelos según la Ec.3.5. PW Op=k (asteriscos rojos). PW Cp=k (cuadrados azules). CW-GOI (diamantes magenta). CW-GCI (cruces oscuras).**

Lo anterior expuesto queda en correspondencia con lo explicado en este capítulo. No obstante, esta alternativa no permite separar entre los distintos modelos de PW, por lo que consideramos que la determinación de la pertenencia de una señal a un modelo basándonos en el criterio de correlación es superior.

### 3.5 Ocurrencia de Outliers

Se ha mostrado en la Figura 3.1 los valores promedio de las correlaciones asumidas presentes para cada modelo. Cada señal pertenece a un modelo si en este obtiene el mayor valor promedio de sus tres correlaciones. Los *outliers* (valores muy bajos) pueden ocurrir sólo si no hay correlaciones grandes (mayores que 0.5) entre las 8 calculadas. Esto puede ocurrir debido a que la variabilidad de alguna señal seleccionada no sea suficientemente grande o que la componente de ruido sea muy grande.

En el modelo PW ocurren (5) *outliers*, los cuales ocurren cuando una de las tres hipótesis de correlación para el modelo no alcanza el umbral de moderado ( $r=0.5$ ). Solamente uno de los *outliers* tiene un valor por debajo de 1/3, debido al fallo de dos correlaciones que no alcanzaron el nivel de moderado.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

- 1- En esta investigación se seleccionaron 114 señales de la base de datos de voz *Saarbrücken* basados en una inspección visual de las señales EGG. Estas fueron escogidas teniendo en cuenta dos aspectos, la variabilidad e irregularidad.
- 2- Se evaluaron cuatro modelos PW con Op constante, PW con Cp constante, CW-GOI y CW-GCI. En los modelos CW el pulso glotal se contrae o expande como un todo. La relación entre ambos pulsos se comporta como una línea recta. En el modelo PW con Op constante la pendiente de la línea referente a la fase de apertura de cualquier par de pulsos glotales es igual a uno. Este modelo usa un tren de impulsos como excitación a la función de transferencia del tracto vocal  $h(t)$ , con esta aproximación todas las fases (Op) diferentes de cero de pulsos glotales son iguales. El modelo PW con la variedad Op constante no implica la inducción de shimmer por jitter.
- 3- Se evaluaron los modelos en función de 8 posibles correlaciones entre fases y duraciones totales. El modelo basado en PW con Op constante emerge como el más adecuado. La fase GCIs debe ser la seleccionada para el procedimiento de síntesis y análisis, ya que no tiene efecto en las alternativas basadas en PW y es más representante en la rama CW. Estos resultados ayudan a modificar lo visto hasta ahora respecto a que los PDAs ciclo a ciclo dejan de ser confiables para valores de jitter por encima del 6%.

**Recomendaciones**

- 1- La síntesis de voces patológicas debe realizarse lo más cercano a la realidad por lo que se recomienda el uso del modelo PW con  $O_p$  constante.
- 2- El modelo CW-GOI no es adecuado para realizar la síntesis de señales.
- 3- La selección de señales a la hora de evaluar modelos debe hacerse de manera muy cuidadosa para así minimizar o eliminar la aparición de outliers.

**REFERENCIAS BIBLIOGRÁFICAS**

- AICHA, A. B. 2018. Voice source modelling using modified LF model with reduced parameters. *Procedia Computer Science*, 126, 423-430.
- AIRAS, M. 2008. TKK Aparat: an environment for voice inverse filtering and parameterization. *Logopedics Phoniatrics Vocology*, 33, 49-64.
- AKANDE, O. O. & MURPHY, P. J. 2005. Estimation of the vocal tract transfer function with application to glottal wave analysis. *Speech Communication*, 46, 15-36.
- ALKU, P. & LAUKKANEN, A.-M. 1998. Estimation of amplitude features of the glottal flow by inverse filtering speech pressure signals. *Speech Communication*, 24, 123-132.
- ALKU, P., MAGI, C. & BÄCKSTRÖM, T. 2009. Glottal inverse filtering with the closed-phase covariance analysis utilizing mathematical constraints in modelling of the vocal tract. *Logopedics Phoniatrics Vocology*, 34, 200-9.
- ANANTHAPADMANABHA, T. V. & YEGNANARAYANA, B. 1979. Epoch Extraction from Linear Prediction Residual for Identification of Closed Glottis Interval. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27, 309-319.
- ANÍBAL J. FERREIRA, J. M. T. A HOLISTIC GLOTTAL PHASE-RELATED FEATURE. *In: DAFX*, ed. Proceedings of the 21 st International Conference on Digital Audio Effects, 2018 Aveiro, Portugal.
- ASOKE KUMAR DATTA, R. S., KAUSHIK BANERJEE AND DIPAK GHOSH 2019. Acoustical Analysis of the Tanpura. *In: TECHNOLOGY*, S. A. C. (ed.).
- BAKEN, J. R. 1992. Electroglottography. *Journal of Voice*, 6.
- BANGAYAN, P., LONG, C., ALWAN, A., KREIMAN, J. & GERRATT, B. R. 1997. Analysis by synthesis of pathological voices using the Klatt synthesizer. *Speech Communication*, 22, 343-368.
- BERRY, D. A. 2001. Mechanism of modal and non-modal phonation. *Journal of Phonetics*, 29, 431-450.
- BOERSMA, P. & VAN HEUVEN, V. 2001. Speak and unSpeak with PRAAT. *Glott International*, 5, 341-347.
- BOUZID, A. A. E. "Glottal Opening Instant Detection from Speech Signals". Proc. of the 12th European Signal Processing Conference, 2004.

- C. MANFREDI, A. G., J. SCHOENTGEN, S. FRAJ, L. BOCCHI, AND P. H. DEJONCKERE 2012. "Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Biomed. Signal Process. Control*, 7.
- CHEN, G., SHUE, Y.-L., K REIMAN, J., ALWAN, A 2012. Estimating the voice source in noise. *Interspeech*.
- D.G. CHILDERS, D. M. H., G. P. MOORE, Y.A. ALSAKA 1986. A model for vocal fold vibratory motion, contact area, and the electroglottogram. *J.Acoust.Soc.Am*.
- DEJONCKERE.H.P, G. A., SCHOENTGEN.J, FRAJ.S, BOCCHI.L, MANFREDI.C 2012. To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Biomedical signal processing and control*, 7.
- DRUGMAN, T., ALKU, P., ALWAN, A. & YEGNANARAYANA, B. 2014. Glottal source processing: From analysis to applications. *Computer Speech and Language*, 28, 1117-1138.
- DRUGMAN, T., DUTOIT, T 2009. Glottal closure and opening instant detection from speech signals. *Interspeech*.
- E. CATALDO , C. S. 2016. Jitter generation in voice signals produced by a two-mass stochastic mechanical model. *Biomedical Signal Processing and Control*, 27, 87-95.
- FABRE, P. 1957. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. *Bulletin de l'Academie Nationale de Medecine*.
- FANT.G 1970. Acoustic Theory of Speech Production. *In: MOUTON, T. H. (ed.)*.
- FANT.G, L. J., LIN.Q 1985. A four-parameter model of glottal flow. *KTH Computer Science and Communication*.
- FARIBORZ ALIPOUR, C. B., DOUGLAS D. COOK, ANDREAS GÖMMEL,ET..AL 2011. Mathematical Models and Numerical Schemes for the Simulation of Human Phonation. *Current Bioinformatic*.
- FERREIRA, A. On the Physiological Validity of the Group Delay Response of All-Pole Vocal Tract Modeling. Audio Engineering Society Convention 145, 2018. Audio Engineering Society.
- FERRER, C. A., HADERLEIN, T., MARYN, Y., DE BODT, M. S. & NÖTH, E. 2018. Collinearity and sample coverage issues in the objective measurement of vocal quality: the case of roughness and breathiness. *Journal of Speech, Language, and Hearing Research*, 61, 1-24.
- FERRER, C. A., TORRES, D., GONZÁLEZ, E., CALVO, J. R. & CASTILLO, E. Effect of different jitter-induced glottal pulse shape changes in periodicity perturbation measures. Sixteenth Annual Conference of the International Speech Communication Association, 2015.
- FLANAGAN, J. L. 1972. Speech Analysis, Synthesis and Perception. *In: SPRINGER-VERLAG (ed.)*.

- FLANAGAN, J. L., ANDISHIZAKA, K. L. 1975. "Synthesis of speech from a dynamical model of the vocal cords and vocal tract". *Bell System Tech. J.*, 54, 484-506.
- FUJISAKI, H., LJUNGQVIST, M. 1986. Proposal and evaluation of models for the glottal source waveform. *ICASSP*.
- FULCHER, L. P., SCHERER, R. C., ZHAI, G. & ZHU, Z. 2006. Analytic Representation of Volume Flow as a Function of Geometry and Pressure in a Static Physical Model of the Glottis. *Journal of Voice*, 20, 489-512.
- G.FANT 1960. Acoustic Theory of Speech Production. *In: MOUTON, T. H. (ed.)*.
- GÓMEZ-VILDA, P., FERNÁNDEZ-BAÍLLO, R., NIETO, A., DÍAZ, F., FERNÁNDEZ-CAMACHO, F. J., RODELLAR, V., ÁLVAREZ, A. & MARTÍNEZ, R. 2007. Evaluation of Voice Pathology Based on the Estimation of Vocal Fold Biomechanical Parameters. *Journal of Voice*, 21, 450-476.
- HENRICH, D. A. C., DOVAL, B., CASTELLENGO, M. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Hal*.
- HOWARD, D. M. 1995. Variation of electroglottographically derived closed quotient for trained and untrained adult female singers. *J. Voice*, 9, 163-172.
- HUIQUN DENG, R. K. W., MICHAEL PETER BEDDOES, MURRAY HODGSON 2006. A New Method for Obtaining Accurate Estimates of Vocal-Tract Filters and Glottal Waves From Vowel Sounds. *AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 14.
- J.M, M. 1999. Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *Acoustical Society of America*.
- J.R. DELLER, J. H. L. H., J.G. PROAKIS 1993. Discrete-Time Processing of Speech Signals. *In: PRESS, I. (ed.)*.
- JANKOWSKI, C., QUATIERI, T., REYNOLDS, D. 1995. Measuring fine structure in speech: Application to speaker identification. *ICASSP*.
- KANE, J. & GOBL, C. 2013. Evaluation of glottal closure instant detection in a range of voice qualities. *Speech Communication*, 55, 295-314.
- KLATT, D. H. 1980a. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.
- KLATT, D. H. 1980b. Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67, 971-995.
- L.R. RABINER, R. W. S. 1978. Digital Processing of Speech Signals. *In: PRENTICE-HALL (ed.)*.
- M. PÜTZER, W. J. B. "Saarbrücken Voice Database."
- MA, Y. K. C., WILLEMS, L. F. 1994. A Frobenius norm approach to glottal closure detection from the speech signal. *IEEE Trans. Speech Audio Processing*.

- MARCELO DE OLIVEIRA ROSAA, J. C. P. 2003. A contribution to simulating a three-dimensional larynx model using the finite element method. *Acoustical Society of America*.
- MILLER, R. L. 1959. Nature of the vocal cord wave. *J Acoust Soc Am*, 31.
- MORA, C. A. C. 2014. Objetividad de la electroglotografía. Aplicaciones clínicas e investigativas en la voz. *LOGOPEDIA, FONIATRÍA y AUDIOLOGÍA*.
- MURPHY, J. P., M. G. K., WALSH, M., COLREAVY, M. 2008. Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals. *Acoustical Society of America*.
- NAYLOR, P. A., KOUNOUEDES, A., GUDNASON, J., BROOKES, M., JAN 2007. Estimation of glottal closure instants in voiced speech using the DYPSA algorithm. *IEEE Trans. Audio, Speech Lang.*
- ROSENBERG, A. E. 1971. Effect of glottal pulse shape on the quality of natural vowels. *the Journal of the Acoustical Society of America*, 49, 583-590.
- ROTHENBERG, M., AND MAHSHIE, J. J. 1988. Monitoring vocal fold abduction through vocal fold contact area. *J. Speech Hear*, 31, 338-351.
- SARA ANASTAPLO, M. P. K. 1988. Synchronized videostroboscopic and electroglottographic examination of glottal opening. *J. Acoust. Soc. Am*, 5.
- SHUE, Y.-L., ALWAN, A. 2010. A new voice source model based on high-speed imaging and its application to voice source estimation. *ICASSP*.
- SMITS, R. & YEGNANARAYANA, B. 1995. Determination of instants of significant excitation in speech using group delay function. *IEEE Transactions on Speech and Audio Processing*, 3, 325-333.
- STRUBE, H. W. 1974. Determination of the instant of glottal closure from the speech wave. *the Journal of the Acoustical Society of America*, 56, 1625-1629.
- THOMAS, M. R. P., GUDNASON, J., NAYLOR, P. A. 2012. Estimation of glottal closing and opening instants in voiced speech using the yaga algorithm. *IEEE Trans. on Audio, Speech & Language Processing*
- TITZE, I. R. 1973. The Human vocal cords: A Mathematical Model. *Phonetica*.
- TITZE, I. R. 1995. *Workshop on acoustic voice analysis: Summary statement*, National Center for Voice and Speech.
- TUAN, V. N., D'ALESSANDRO, C. Robust glottal closure detection using the wavelet transform. *European Conf. Speech Processing, Technology*, 1999 Budapest.
- VELDHUIS, R. 1998. A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation. *Acoustical Society of America*.
- WENDAHL, W. R. 1966. Some Parameters of Auditory Roughness. *Folia Phoniatica*

## ANEXOS

Segmentos escogidos de la base de datos de *Saarbrücken*

Señal	Segmento Inicio	Segmento Fin	Señal	Segmento Inicio	Segmento Fin
'0110'	1	25000	'1675'	23000	3801
'0143'	74000	13036	'1686'	1	34000
'0143'	39000	45000	'1752'	1	27285
'0360'	2000	8000	'1769'	27000	32500
'0670'	23000	36000	'1776'	12000	20000
'0715'	23000	16451	'1795'	1	81962
'0715'	1	13000	'1798'	15000	19500
'0722'	5600	6300	'1802'	53000	60000
'0850'	30000	40000	'1803'	16000	21000
'0855'	6000	24700	'1828'	44500	48000
'0868'	48900	4384	'1867'	1	6800
'0868'	13000	17000	'1887'	51000	60000
'0871'	89000	91000	'1892'	26000	39000
'0875'	3000	23000	'1898'	72000	14937
'0891'	18000	58800	'1906'	48000	65000
'0891'	1	13000	'1913'	22000	26500
'0892'	48000	57500	'1914'	2000	11000
'0893'	8000	15000	'1931'	10000	22797
'0896'	1	25000	'1946'	22000	30000
'0899'	53000	64000	'1966'	6000	16000
'0914'	6700	13000	'1967'	4000	20000
'0916'	4000	19000	'1973'	40000	52000
'0935'	19000	22000	'1988'	36000	42000
'0938'	58000	14117	'1998'	45000	62000

'0938'	11000	41000	'2002'	36000	43000
'1045'	1	29944	'2005'	62000	5983
'1046'	48000	56000	'2014'	28000	35372
'1048'	47000	53000	'2025'	42000	24236
'1117'	72000	19707	'2028'	1	45609
'1117'	30000	61000	'2063'	80000	83000
'1120'	11000	32304	'2085'	21500	26500
'1157'	12000	20000	'2097'	4000	20000
'1186'	65000	85000	'2119'	28000	1
'1189'	13000	26000	'2121'	1000	4000
'1237'	20000	15947	'2135'	41000	11243
'1253'	20000	27000	'2194'	5000	8500
'1264'	40000	67000	'2213'	1	8000
'1268'	22000	32000	'2229'	1	61228
'1285'	16000	32000	'2241'	1500	6000
'1301'	3000	16000	'2246'	22000	28000
'1322'	53000	70000	'2265'	18000	36000
'1382'	63000	5443	'2298'	1	27644
'1382'	32000	35000	'2305'	1	37223
'1382'	4000	8000	'2315'	52000	60000
'1400'	11000	15000	'2325'	8000	15000
'1404'	23000	27000	'2329'	19000	6388
'1408'	58000	61000	'2398'	90500	94000
'1421'	1	26094	'2410'	1	9000
'1445'	29200	32500	'2415'	1	13000
'1453'	10000	26000	'2429'	2000	47724
'1455'	4000	8500	'2430'	1	64289
'1463'	42000	6957	'2457'	23000	34659
'1478'	15500	21000	'2468'	20000	28000
'1549'	53000	60000	'2488'	48000	62000
'1552'	20000	30000	'2493'	22000	28000
'1562'	6500	11500	'2497'	25000	38500
'1571'	15000	30000	'2502'	33000	43000
'1576'	53000	70000	'2504'	17000	32000
'1610'	42000	11492	'2515'	48000	52000
'1613'	57000	63000	'2537'	1	25000
'1639'	1	26000	'2544'	1	29932

---

'1656'	70000	78000	'2565'	6500	11500
'1672'	59000	63000	'2571'	35000	29640
			'2588'	1	16000