



UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

**ANÁLISIS DEL ZAPOTECO MEDIANTE
TÉCNICAS DE RECONOCIMIENTO DE
PATRONES.**

Tesis que presenta:

Ing. Gabriela Oliva Juárez

Para obtener el grado de

Maestra en Ciencias y Tecnologías de la Información

Asesora (s): M. en I. Fabiola M. Martínez Licona
M. en C. Alma E. Martínez Licona

Jurado calificador:

Presidente: Dr. Irineo López Cruz

Secretaria: M. en C. Alma E. Martínez Licona

Vocal: Dr. John Goddard Close

México, D.F., Enero 2014

Contenido.

Resumen.....	4
1. Introducción.....	5
2. Antecedentes.....	12
3. Objetivos.....	14
3.1. Objetivo General.....	14
3.2. Objetivos Particulares.....	14
4. Metodología.....	15
4.1. Base de datos.....	15
4.2. Transcripción Fonética y Segmentación.....	17
4.3. Determinación de Fonemas.....	24
4.4. Extracción de Características.....	29
4.5. Análisis.....	31
4.6. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).	33
5. Resultado.....	39
5.1. Segmentación.....	39
5.2. Fonemas.....	40
5.3. Reconocedor DTW.....	41
5.4. Determinación de Fonemas.....	42
5.5. Extracción de características.....	45
5.6. Análisis.....	46
5.7. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).	55
6. Discusión de Resultados.....	59
7. Conclusiones y Trabajo Futuro.....	64
Anexo A. Variantes de la lengua zapoteca, su estándar ISO639.3 y escala EGIDS.....	65
Anexo B. Zapoteco de Veracruz.....	68
Anexo C. Playa Vicente y Arenal Santa Ana.....	69
Anexo D. Lista Swadesh.....	74
Anexo E. Transcripción Fonética de la Base de datos de Choapan de Veracruz.....	78
Anexo F. Pasos de la Segmentación HTK.....	81
Anexo G. Alineamiento Temporal Dinámico (Dynamic Time Warping, DTW).....	97
Anexo H. Extracción de características en Praat.....	99
Anexo I. Perceptrón Multicapa (MLP) y Máquinas de Soporte Vectorial (MSV).....	102
Anexo J. Palabras utilizadas para determinar el símbolo fonético de Ě.....	105
Anexo K. Determinar Fonemas.....	108
Anexo L. Resultados de los Métodos de Reconocimiento de Patrones.....	119
Anexo M. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).....	124
Anexo N. Artículo.....	137
8. Referencias.....	145

Resumen.

La lengua zapoteca, así como muchas otras lenguas prehispánicas en México, está en peligro de desaparecer por varias razones. Entre éstas destacan la falta de uso por parte de la población más joven que prefieren hablar español en lugar de su lengua materna, y la desaparición de los hablantes de mayor edad. Una característica importante del zapoteca es que esta lengua posee un número considerable de variantes, las cuales no suelen tener un nivel de inteligibilidad alto entre ellas, ocasionando que algunas prevalezcan y otras se vayan extinguiendo con mayor celeridad.

En este trabajo se presenta un análisis de la variante zapoteca de Choapam a través de las señales de habla; se obtuvo una lista de palabras aisladas a través de la grabación de un hablante de esta variante así como la grabación de un texto por parte de otro hablante. Posteriormente se realizó una segmentación fonética de las palabras para identificar los sonidos sonoros y sordos; dado que el zapoteca tiene una característica tonal sobresaliente, el análisis se enfocó en las vocales. Se obtuvieron las características de duración, frecuencia fundamental, intensidad y formantes tanto del conjunto de vocales en general como de vocales en sus tres diferentes realizaciones: sencilla, cortada y quebrada. El proyecto se enfocó en abordar dos problemas: la clasificación de las vocales y la identificación del sonido vocálico *ë*, el cual no se encuentra dentro del alfabeto del español y presenta una similitud a la vocal *e*.

Los resultados muestran que el problema de clasificación se encuentra precisamente entre las vocales *e* y *ë* debido a su similitud acústica, y entre las realizaciones de todas las vocales; es importante hacer notar que el cambio de vocal provoca un cambio de significado en la palabra, lo cual impacta negativamente la inteligibilidad entre los hablantes. Con respecto a la identificación del sonido *ë*, se aplicó la técnica de alineación temporal dinámica (dynamic time warping DTW) obteniendo el mejor resultado con sonidos del alemán y del sueco. Cabe mencionar que estos resultados están dependientes de la plataforma en la que se obtuvieron (Praat para el alemán y Matlab para el sueco).

Este proyecto aborda un primer enfoque para el análisis de lenguas bajo amenaza de extinción; se espera que el uso de esta información contribuya a nuevos esfuerzos dirigidos a ayudar a preservar el idioma.

1. Introducción.

Los seres humanos necesitan comunicarse para llegar a entendimientos o acuerdos que beneficien el progreso de las comunidades, pueblos e incluso naciones que éstos habitan, para lo cual utilizan un conjunto de símbolos, gestos, sonidos y señas que conforman lo que se denomina lengua [1, 2]. Una importante característica de ésta es que no es constante en el tiempo, lo que conlleva al desuso y por lo tanto a su extinción. Ejemplo de esto ocurrió con las lenguas Eudave, Algonquino cristiano, Cholti, Cochimi, Ópata, Massachusset, Tapaculteco, Mangué, Chapaneco y Subtiada, todas ellas habladas en el Continente Americano [3]. Por otro lado, los cambios en una lengua pueden dar lugar a una gran diversidad cultural en el mundo; en este sentido, actualmente se cuenta con 7105 lenguas vivas que se clasifican de acuerdo a su familia lingüística, subgrupos, macrolenguas, etc.

La familia lingüística contiene a las lenguas o variantes lingüísticas que presentan semejanzas estructurales y léxicas que se deben a un origen histórico común, además de que el nombre de la familia depende de las lenguas que la integran. Actualmente en México conviven once familias lingüísticas que son: Álgica, Yuto-nahua, Cochimí – yumana, Seri, Oto-mangué, Maya, Totonaco – tepehua, Tarasca, Mixe-zoque, Chontal de Oaxaca y Huave [3], y cuya distribución geográfica se muestra en la figura 1.1.



Fig.1.1. Familias lingüísticas que se encuentran a lo largo del territorio mexicano. [3]

Una familia lingüística está integrada por una o más variantes las cuales se definen como aquellas lenguas que se hablan en una pequeña comunidad delimitada geográficamente [3]. Las variantes lingüísticas suelen clasificarse dentro de grupos y subgrupos; algunos de ellos se denominan macrolenguas cuando las variantes que lo integran se encuentran estrechamente relacionadas ya sea porque sus hablantes comparten una herencia e identidad, además de contar con un sistema de escritura y literatura en común [4]. Los hablantes de diferentes variantes que pertenecen a una misma Macrolengua ven afectada la inteligibilidad de los mensajes que intercambian, la cual puede ir disminuyendo hasta ser nula y por lo tanto el mensaje no se comprenda, o tan alta que el mensaje se entienda en su totalidad.

Dentro de las once familias lingüísticas que se encuentran en México, la más grande y diversificada es la familia Oto-mangue ya que las lenguas que la integran son habladas desde el estado de San Luis Potosí con la lengua Pame, hasta el sur en el estado de Oaxaca con la lengua zapoteca [3]. Todas las lenguas que conforman esta familia lingüística se encuentran en el diagrama de árbol de la figura 1.2 donde, mediante el número de la derecha se indica la cantidad de variantes lingüísticas que posee cada una de ellas [5], y en el mapa de la figura 1.3 donde se presenta su distribución geográfica en el territorio mexicano. Una de las cualidades de las lenguas que pertenecen a la familia Oto-mangue es la tonalidad [6]; los hablantes de estas lenguas hacen cambios en la frecuencia con que vibran las cuerdas vocales con el fin de distinguir palabras iguales en ortografía pero con diferente significado.

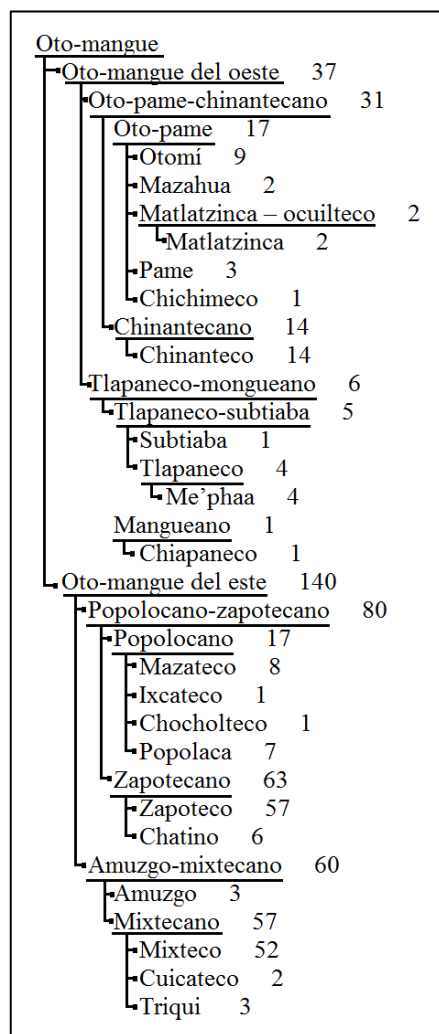


Fig.1.2. Diagrama de árbol de la Familia Oto-mangue. [4]



Fig.1.3. Mapa que muestra la distribución de la familia lengua Oto-mangue en el territorio mexicano [3].

En el diagrama de árbol para la familia Oto-mangue encontramos que el subgrupo Zapotecano se compone de 63 variantes lingüísticas, 6 pertenecen a la lengua chatino y 57 son de la lengua zapoteca, esta última definida como una macrolengua [4]. La cantidad de variantes del zapoteco puede ser atribuida a las condiciones políticas o geográficas del terreno además de la falta de disponibilidad por parte de los hablantes para establecer una sola lengua, ya que éstos sólo la utilizan en la comunidad en que habitan y no para entablar una comunicación con hablantes de otras comunidades [7].

Todas las lenguas reconocidas por el Instituto Lingüístico de Verano (SIL por sus siglas en inglés) se encuentran identificadas por medio del estándar ISO639.3 que consiste de un código de tres letras. Por lo tanto, la macrolengua zapoteca junto con sus 57 variantes se identifican mediante el código “zap”, mientras que la variante Zoogocho con “zpg”, la variante de Choapan* con “zpc”, etc., [2] la lista completa del código para cada variante del zapoteca se encuentra en el Anexo A.

La lengua zapoteca se habla en los estados de Oaxaca, Veracruz, Edo. de México e incluso en Estados Unidos. Según el censo realizado por el INEGI en el 2010, México cuenta con una población de 6 millones 695 mil 228 personas cuya edad es igual o mayor a 5 años que hablan alguna lengua indígena y de los cuales 450 mil 419 son hablantes de alguna variante de la lengua zapoteca [8]. Aunque este censo indica que la población de hablantes de la lengua zapoteca aumentó, esto no aplica a todas las variantes que la integran ya que algunas se encuentran en un nivel bajo de desarrollo de la lengua, y por lo tanto al borde de la extinción, según la escala EGIDS (Expanded Graded Intergerational Disruption Scale) [9].

La EGIDS es una escala multidimensional que analiza los niveles de vitalidad de una lengua, es decir, mide el nivel de interrupción en el uso de ésta. La escala consiste de 13 niveles como se indica en la tabla 1.1; el nivel más alto en ella es el 10, el cual representa el mayor nivel de interrupción intergeneracional, lo que conlleva a que la lengua quede en un desuso completo ya que los hablantes no se identifican con ningún aspecto de ella. Por otro lado el nivel más bajo es el 0, el cual indica que una lengua es utilizada internacionalmente [5].

* La variante de Choapan también es denominada como la variante de Choápam.

Tabla.1.1. Estados y niveles de las lenguas según la escala EGIDS.

Nivel	Estado
0	Internacional
1	Nacional
2	Provincial
3	Amplia Comunicación
4	Educacional
5	Desarrollo
6a	Vigoroso
6b	Amenazado
7	Cambiando
8a	Moribundo
8b	Cerca de la Extinción
9	Latente
10	Extinto

La escala EGIDS se grafica utilizando la relación entre el nivel de desarrollo o peligro (eje horizontal) contra el valor logarítmico de la población (eje vertical). La grafica EGIDS coloca un punto por cada una de las lenguas del mundo con el fin de comparar la situación de ésta con respecto al resto de lenguas, de tal forma que aquella que se encuentra en un nivel más extendido, su punto se localiza más cerca de la esquina superior izquierda y mientras más se acerque a la parte inferior derecha de la gráfica, la lengua corre más riesgo de estar en peligro de extinción [5]

Las 57 variantes lingüísticas del zapoteco se encuentran en diferentes niveles dentro de la escala EGIDS, 20 de ellas se encuentran en el estado de “desarrollo” equivalente al quinto nivel de la escala que se caracteriza por contar con una literatura estandarizada utilizada por unas cuantas personas, lo que causa que su uso no sea tan extenso o sostenible. El resto de variantes zapotecas se ubican dentro de la escala EGIDS en una situación de mayor riesgo, es decir, se distribuyen en niveles mayores al quinto, de la siguiente forma (ver Anexo A): quince están en el nivel “vigoroso” (6a), diez tienen un estado amenazado (6b), siete están cambiando (7), tres están en un estado moribundo (8a) y solo dos están cerca de la extinción (8b) [5].

La variante de Choapan se encuentra dentro del estado moribundo (8a) en la escala EGIDS, es decir, esta lengua es usada solo por personas mayores o de la tercera edad [5]. El punto resaltado en la figura 1.4 indica la situación de esta variante lo que permite comparar su estado contra el resto de las lenguas del mundo. La variante de Choapan cuenta con una población de 12000 hablantes según la SIL en sus registros del 2007, se habla principalmente en el distrito de Choapan del estado de Oaxaca así como en los municipios de Playa Vicente y Santiago Sochiapan del estado de Veracruz (para mayor información sobre el zapoteco que se habla en éste estado se puede consultar el Anexo B). La variante de Choapan presenta una inteligibilidad del 60% con la variante Zoogocho [4].

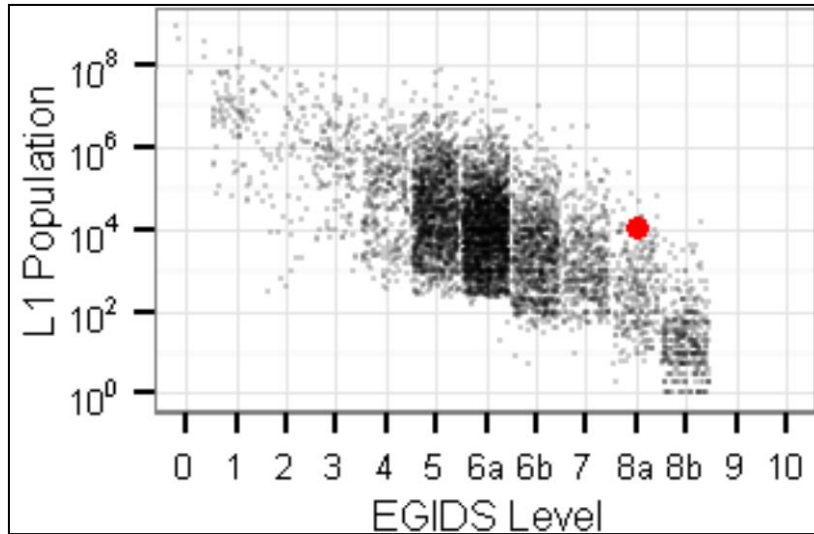


Fig.1.4. Gráfica EGIDS para la variante de Choapan [5].

La lengua puede analizarse a partir de sus sonidos elementales, o fonemas, los cuales se representan con su alfabeto. El alfabeto de la variante zapoteca de Choapan se asemeja al utilizado en el español, con la diferencia de que el zapoteco tiene una mayor cantidad de letras [10]. En lo que respecta a las vocales, éstas son 6: cinco similares a las utilizadas en el español (A, E, I, O, U) más una sexta (Ë), todas ellas con tres tipos de realización:

- Las sencillas son similares a las empleadas en el español.
- Las cortadas terminan de forma abrupta mediante un corte de voz.
- Las quebradas (interrumpidas) se forman a partir de la unión de dos vocales iguales con un pequeño corte de voz entre ellas.

De acuerdo a lo anterior, el total de vocales empleadas en la variante de Choapan asciende a 18 donde la sintaxis utilizada para cada una de ellas se encuentra en la tabla 1.2 y la señal como un espectrograma* de las mismas se muestra en la figura 1.5. Además de la gran cantidad de vocales del zapoteco, también se utilizan diptongos que a diferencia de las vocales quebradas éstos no realizan un corte de voz en medio de su realización y pueden estar formados por letras diferentes, por ejemplo: AI, UA, UE, UI, etc.

Tabla.1.2. Vocales empleadas en el zapoteco de Choapan en sus tres diferentes realizaciones.

Vocales	A	E	Ë	I	O	U
Ejecuciones						
Sencillas	A	E	Ë	I	O	U
Cortadas	A'	E'	Ë'	I'	O'	U'
Quebradas	A'A	E'E	Ë'Ë	I'I	O'O	U'U

*Espectrograma. Es una herramienta que permite observar la señal del habla de forma tridimensional, donde observamos proyectados sobre el tiempo, la intensidad de la señal en diferentes bandas de frecuencia. [11]

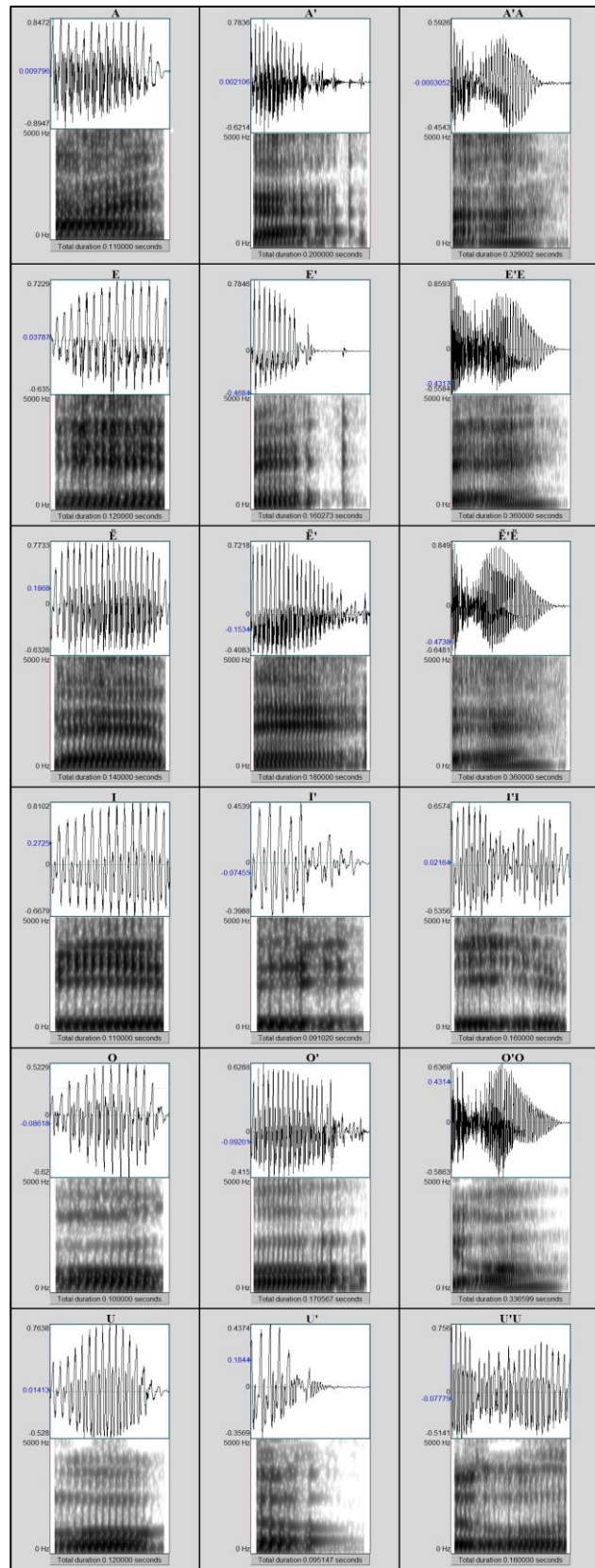


Fig.1.5. Señal y espectrograma de las vocales de la variante de Choapan.

Las consonantes de la variante de Choapan están integradas por 23 letras, las cuales se muestran en la tabla 1.3, en su mayoría parecidas al español a excepción de las letras que no se utilizan en él que son: DZ, DY, TZ y Ž. Algunos puntos a resaltar sobre las consonantes son los siguientes:

- Se pueden clasificar como oclusivas, fricativas, africadas, vibrantes, etc. La señal y su espectrograma para algunos de estos tipos de constantes se muestra en la figura 1.6.
- La letra “R” tiene dos usos al igual que en el español (“R” o “RR”).
- El alfabeto no utiliza las consonantes “Ñ” y “F”.
- La letra “K” es empleada sólo en palabras que provienen del español.
- La “QU” sólo antecede a las vocales I ó E.

Tabla.1.3. Consonantes del alfabeto variante de Choapan.

Consonantes																						
B	C	QU	CH	D	DZ	DY	G	HU	J	K	L	M	N	P	R	S	T	TZ	X	Y	Z	Ž

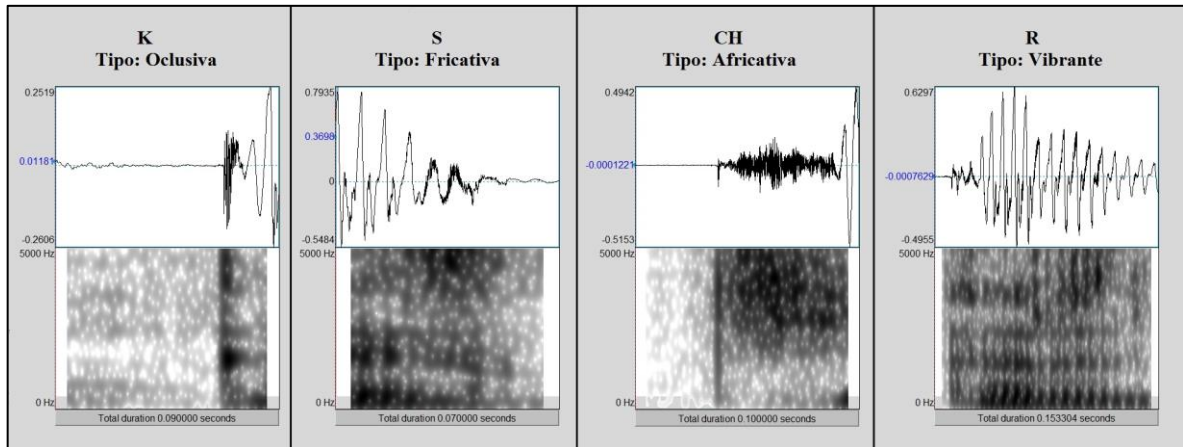


Fig.1.6. Señal y espectrograma para algunos tipos de consonantes de la variante de Choapan.

Este trabajo se enfocó en analizar los sonidos de las vocales para la variante del zapoteca de Choapan a partir del estudio de las características acústicas de sus fonemas.

2. Antecedentes.

Actualmente el Instituto Lingüístico de Verano (SIL por sus siglas en inglés) tiene registradas 7105 lenguas [5]. Algunas de ellas sólo se conocen por su nombre ya que no se posee un estudio sobre los sonidos, gestos o símbolos que los hablantes utilizan, y al tener éstos contacto con otros individuos que hablan una lengua diferente, el evento provoca que con el tiempo se pueda crear una nueva variante de la lengua, una nueva lengua o simplemente que los hablantes dejen de usarla. Es el desuso una de las principales causas que da lugar a la necesidad de conocer y preservar los sonidos de la gran mayoría de las lenguas mediante trabajos que aborden su análisis; los tipos de análisis más comunes consisten en extraer características esenciales de los sonidos como los valores de las formantes, la intensidad, la duración o la frecuencia fundamental (f_0), los cuales pueden ser de utilidad para realizar estudios posteriores enfocados al reconocimiento o síntesis.

La frecuencia fundamental es una característica que resulta de obtener el valor de la vibración de las cuerdas vocales durante la emisión de un sonido y consiste en una forma de medir el tono. El tono permite la diferenciación entre palabras que presentan igual ortografía [12] ocasionando que las lenguas que se basan en los tonos tengan un alto grado de dificultad tanto para realizarse como para aprenderse [13]. La lengua tonal se distingue por realizar cambios en el tono con el fin de hacer contraste en las palabras como en el caso del chino mandarín, el cantones, el tailandés o el yorúba. La importancia del tono en estas lenguas se ha evidenciado de los estudios realizados en lenguas como el chino o el tailandés, los cuales se enfocaron en su identificación, encontrando un alto grado de dificultad en la tarea de reconocimiento, teniendo presente que un error en éste puede degradar el grado de inteligibilidad entre los hablantes [14, 15, 16]. En México, el ejemplo más extendido de lengua tonal recae en la familia lingüística otomangue, la cual incluye las lenguas chinanteca, mixteca y zapoteca; esta última, al contar con 57 variantes, presenta un problema de inteligibilidad importante [17]. Este problema en el entendimiento del mensaje puede ser provocado por una inexistente distribución de tonos en la lengua, dado que cada variante utiliza diferentes tipos y cantidades de éstos. En consecuencia existen variantes del zapoteca que presenta tres tonos (alto, medio y bajo) mientras que otras pueden llegar hasta cinco (los tres citados adicionando los tonos ascendente y descendente). Por ejemplo, si se consideran dos pueblos que tienen el mismo origen y que actualmente, debido a su proceso histórico se encuentran separados por una distancia muy corta, se podría pensar que ambos utilizarían el mismo conjunto de tonos. Sin embargo esto no es así, como en el caso de los poblados de San Pedro y San Juan Metepec de Miahuatlán estado de Oaxaca, donde el primero se encuentra geográficamente más alejado de la ciudad y por lo tanto utiliza un sistema de cinco tonos mientras que el segundo sólo utiliza un sistema de cuatro tonos [18].

La diversidad de variantes que posee el zapoteco demanda la atención del análisis tonal con mayor detalle como en el caso del estudio realizado en la variante hablada en el poblado de San Pablo Güilá donde se propuso encontrar la unidad prosódica portadora del tono realizando pruebas entre los cinco diferentes tipos de rima, los seis timbres y

cuatro tonos para concluir que ésta es la rima [19]. Por otro lado para algunos la lengua zapoteca no es del todo tonal sino más bien una lengua con acento tonal, tal como lo demuestra el análisis que se realizó a los tres tonos y a los cuatro tipos de vocales que se utilizan en la variante del poblado de San Baltasar [20]. Otro punto a resaltar con respecto al tono, y tomando al mismo tiempo características adicionales como el volumen y la intensidad que son utilizados dentro del zapoteco y otras lenguas como la turca o la farsi (lengua iraní), es que las tres se encuentran libres de contexto, es decir, un hablante no eleva el tono, el volumen o la intensidad para hacer referencia a objetos de gran tamaño, fuertes y altos o viceversa [21].

Además del tono el análisis de una lengua se puede enriquecer al incorporar otros tipos de características como las formantes, que se definen como el resultado de las resonancias producidas a lo largo del tracto vocal. Su utilización para identificar los sonidos sonoros ha sido ampliamente utilizada y en algunos casos, como el español, con una alta precisión. Algoritmos de aprendizaje maquina pueden utilizarse para la identificación de vocales en lenguas tonales [22], para algunos otros la identificación de las vocales puede obtenerse a partir del análisis de su duración [23] mientras que el uso del espectrograma también se ha ocupado para el análisis de lenguas tonales como en el caso del Sábila [24].

El análisis de una lengua puede obedecer diferentes objetivos. Para el caso de las lenguas prehispánicas puede ser para reconocimiento automático con la finalidad de utilizar aplicaciones [25], para análisis primario a partir de las canciones [26] o para su conocimiento y preservación antes de que tenga una alta probabilidad de desaparecer [27]. Algunas de las variantes de la lengua zapoteca se encuentran en esta situación, tal es el caso de la variante de Choapam el cual se encuentra en gran peligro de desaparecer; con el análisis a partir de la señal de voz obtenida de un hablante de esta variante se tiene la expectativa de obtener información relevante que permita la caracterización y que a su vez de lugar al modelado de la lengua.

3. Objetivos.

3.1. Objetivo General.

Analizar los sonidos de las vocales para la variante de Choapan de la lengua zapoteca desde la perspectiva de las tecnologías del habla.

3.2. Objetivos específicos:

- I. Desarrollar una base de datos en zapoteco para el análisis de la señal de habla.
- II. Seleccionar los componentes acústicos y frecuenciales de la señal del habla en zapoteco para el análisis
- III. Desarrollar el análisis del zapoteco como señal del habla a partir de los componentes seleccionados.
 - a. Descripción estadística de vocales e identificación de los problemas que presenta el lenguaje.

4. Metodología.

La metodología utilizada en el proyecto consiste en las acciones mostradas en la figura 4.1. Como primer paso se generó una base de datos en la variante de Choapam de la lengua zapoteca, seguidamente se efectuó la transcripción y segmentación de la misma a partir de la referencia dada por un alfabeto fonético basado en SAMPA (Alfabeto Fonético Legible por Ordenador). El análisis de la lengua se enfocó en los sonidos vocálicos por lo que se conformaron los conjuntos de vocales de acuerdo a la gramática del zapoteca de Choapam para posteriormente extraer las características acústicas más relevantes de los fonemas vocálicos con la finalidad de realizar una clasificación de los mismos. A continuación se explicarán cada una de las partes de esta metodología.

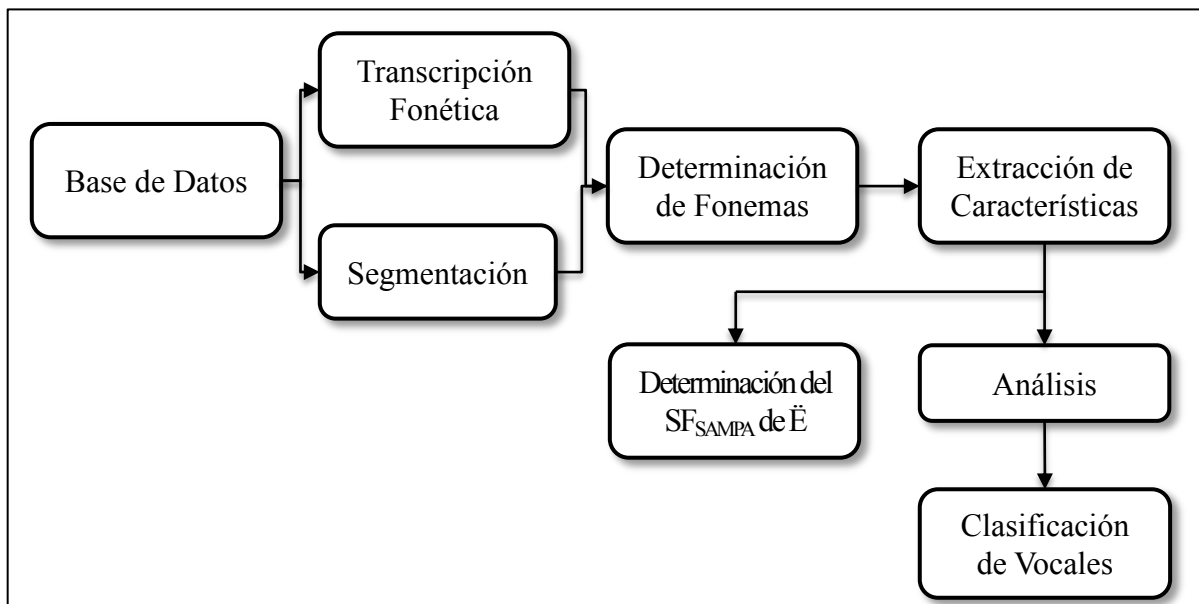


Figura 4.1. Metodología para el análisis de la lengua zapoteca.

4.1. Base de Datos.

La base de datos consiste en dos conjuntos de grabaciones de audios que pertenecen a la variante de Choapan de la lengua zapoteca, con código ISO639.3: “zpc”. Esta variante se habla principalmente en el distrito de Choapan del estado de Oaxaca y también en comunidades de los municipios de Playa Vicente y Sochiapan del estado de Veracruz; según la SIL (Summer Institute Linguistics) esta variante se encuentra en un estado moribundo lo cual significa que solamente la hablan de manera regular los adultos mayores, mientras que las personas jóvenes han dejado de usarla y la han remplazado por el español. El primer conjunto abarca la grabación de unas señales de voz pertenecientes específicamente a la comunidad de Playa Vicente, mientras que el segundo corresponde a un fragmento de la Biblia hablada en la misma variante, el cual se obtuvo de internet; ambos conjuntos se describen a continuación.

Conjunto 1: Base de datos de la variante de Choapan del estado de Veracruz.

Para la elaboración de este conjunto se contó con la colaboración de un hablante de la zona de Playa Vicente, el cual presenta las siguientes características:

Datos del Hablante.

Los audios fueron grabados por una hablante femenina de 21 años originaria de la Comunidad Arenal Santa Ana, municipio de Playa Vicente del estado de Veracruz, aunque actualmente radica en el área metropolitana del Distrito Federal. Esta situación hace que la hablante sea considerada como bilingüe ya que habla tanto zapoteco como español, siendo este último es el que más utiliza. En el Anexo C se encuentra más información sobre el municipio y la comunidad donde es originaria la hablante.

Corpus.

Se grabaron 188 palabras en zapoteco que pertenecen a la lista Swadesh. Esta lista consiste en un conjunto de palabras consideradas comunes en todas las lenguas, la cual incluye adverbios, verbos, artículos, sustantivos, etc. Una descripción de las listas Swadesh así como la lista completa que se empleó se muestra en el Anexo D.

Características de la grabación.

Cada palabra se grabó de forma aislada, es decir, realizando pausas antes y después de su grabación, haciendo dos repeticiones por cada una, con excepción de las palabras: *pie*, *grande* y *ancho* que sólo cuentan con una repetición*. En total la base de datos contiene 373 audios que se almacenaron con la sintaxis <Palabra><Numero> donde

- <Palabra> significado en español del audio grabado en zapoteco.
- <Número> indica la repetición en la que fue grabada la palabra, sólo puede tener los valores 1 o 2.

Los audios se almacenaron en formato “wav” a una frecuencia de muestreo de 16kHz utilizando un micrófono Microsoft modelo Lite Chat Lx-3000. La grabación fue realizada en las instalaciones de la Universidad Autónoma Metropolitana unidad Iztapalapa (UAM-I), minimizando distorsiones (ruido ambiental) que se pudiera presentar.

*Se cuenta con una repetición debido a que no se pudo contactar al hablante para completar la grabación de éstas palabras.

Conjunto 2: Base de datos de la variante del Distrito de Choapan.

Este conjunto se conformó con las grabaciones obtenidas de internet de textos leídos por un hablante masculino [28].

Corpus.

La Biblia es uno de los textos que han sido traducidos a muchas lenguas, incluyendo varias lenguas indígenas mexicanas, de tal forma que se puede tener acceso tanto al texto como al audio en formato mp3 de ciertas partes. El evangelio según San Juan es el libro que consta de 21 capítulos, de los cuales se tomaron los primeros cinco en una versión grabada en zapoteco del Distrito de Choapan. Este libro presenta versículos de diferente tamaño por lo que éstos se segmentaron de tal forma que se obtuvieran enunciados cortos con un máximo de 16 palabras por cada uno. La sintaxis final para los versículos segmentados está dada por `cap<capitulo>parrrf<párrafo><xy>` donde:

- `<capitulo>` proporciona el número de capítulo al cual pertenece el segmento
- `<párrafo>` es la identificación del párrafo, primero con los dígitos del 1 al 9 y en caso de que se necesitara identificar más enunciados se utilizaron las letras mayúsculas del abecedario
- `<xy>` es una combinación de dos letras minúsculas que indica la sección dentro del párrafo; *x* identifica que se trata del mismo versículo mientras que *y* indica las diferentes partes en las que se dividió dicho versículo.

Con la segmentación se obtuvieron 619 archivos en total en un formato .wav a una frecuencia de muestreo de 44.1 KHz utilizando el programa Audacity [29].

4.2. Transcripción Fonética y Segmentación.

El segundo paso de la metodología corresponde a la Transcripción Fonética y Segmentación, el cual consiste de dos tareas que se realizaron al mismo tiempo ya que al segmentar la base de datos con el fin de extraer y agrupar los sonidos del alfabeto de Choapan se les puede asignar a éstos un símbolo que los represente. Para la base de datos de la variante de Choapan del Estado de Veracruz se tomó la Lista Swadesh completa como aparece en las tablas del Anexo D, sin tomar en cuenta si la ortografía de las palabras es correcta. En el caso de la base de datos de la variante del distrito de Choapan (capítulos de la Biblia) no se presenta este problema ya que existe el texto escrito. A continuación se explicara cada una de las partes del segundo paso de la metodología.

Transcripción Fonética.

La Transcripción Fonética consiste en la representación de cada uno de los sonidos de una lengua empleando algún símbolo o signo gráfico que pueda provenir de un alfabeto fonético [30, 31]. En esta parte de la metodología se buscó que la simbología se pudiera representar mediante elementos compatibles en computadora, por lo que la mayoría de los símbolos utilizados surgieron del alfabeto fonético SAMPA [32]. Se utilizó como base la versión de SAMPA para el español por las semejanzas que presenta con el alfabeto zapoteca de Choapan, en el caso de los elementos de este alfabeto que no existen en el español (DX, DY, TZ, Ë, las realizaciones cortadas y quebradas de las vocales) se utilizan símbolos que no se encuentran en el alfabeto SAMPA para

representarlos. La transcripción fonética obtenida se utilizó para realizar la segmentación fonética de las señales de voz usados para el análisis.

Segmentación.

El proceso de segmentación consiste en extraer cada uno de los fonemas (de vocales o consonantes) que integran la base de datos y agruparlos en carpetas de acuerdo al tipo de sonido que los representa; la segmentación se puede realizar de dos formas: manual y automática. En la forma manual el individuo que realiza la segmentación decide los límites de los fonemas a partir de la percepción de las características visibles y acústicas de la señal del habla. En cambio, la forma automática consiste en implementar una técnica o algoritmo que determine los límites entre cada uno de los fonemas y sus adyacentes con el fin de evitar los posibles errores de apreciación. Algunos ejemplos de estas técnicas son: Convez Hull, Spectral Variation Function (SVF) [33], el uso del contorno de energía [34], las relaciones entre picos y valles de la señal [35], etc. En este proyecto la segmentación automática se realizó mediante Modelos Ocultos de Markov (HMM) mediante el software HTK (Hidden Markov Model Toolkit) [36].

Segmentación Automática.

La segmentación automática se implementó en el HTK que permite el modelado de HMM [37]. El proceso de segmentación se dividió en tres etapas como se muestra en el diagrama de la figura 4.2, las cuales son:

1. Transcripción Preliminar
2. Segmentación Automática HTK
3. Extracción de Fonemas

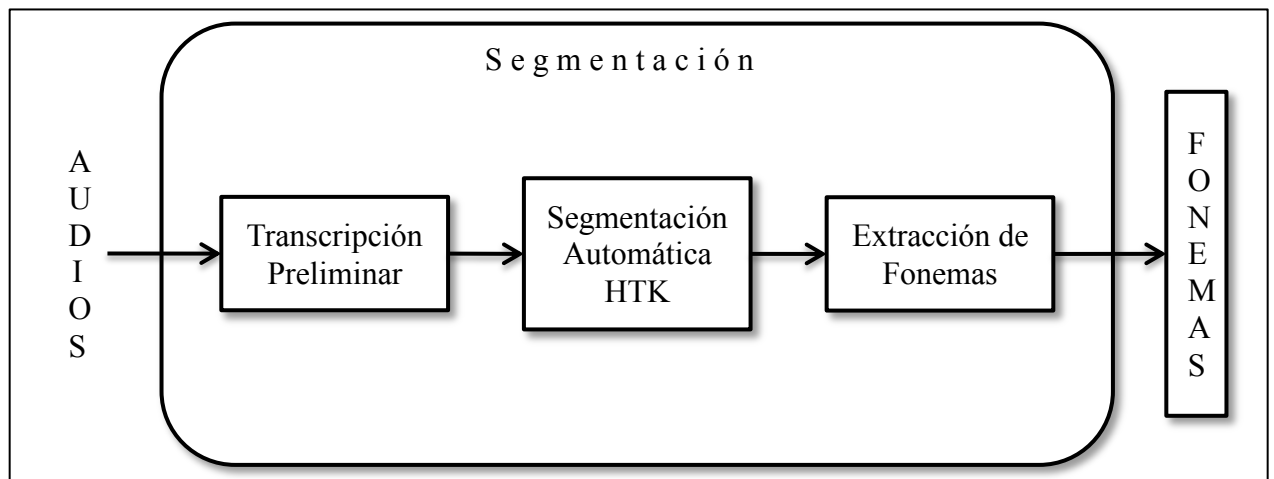


Figura 4.2 Diagrama del proceso de segmentación.

Transcripción Preliminar.

La Transcripción Preliminar consiste en asignar un símbolo a cada uno de los posibles sonidos del zapoteco, tomando en cuenta el alfabeto que se utiliza en esta variante. Se pueden emplear mayúsculas y minúsculas ya que HTK realiza la diferencia entre ambas, lo que nos ayuda a contar con una mayor cantidad de símbolos para

representar los sonidos del alfabeto de la variante de Choapan [10]. En la figura 4.3 se muestran los elementos del alfabeto de la variante de estudio que consta de 6 vocales con 3 realizaciones cada una, y 23 consonantes.

Ejecuciones \ Vocales		Vocales					
		A	E	Ě	I	O	U
Sencillas		A	E	Ě	I	O	U
Cortadas		A'	E'	Ě'	I'	O'	U'
Quebradas		A'A	E'E	Ě'Ě	II	O'O	U'U

Consonantes																						
B	C	QU	CH	D	DZ	DY	G	HU	J	K	L	M	N	P	R	S	T	TZ	X	Y	Z	Ž

Figura 4.3. Alfabeto de la variante de Choapan.

Transcripción Vocales.

La transcripción de las vocales asignó uno o dos símbolos; el primero indica el tipo de vocal y el segundo el tipo de realización aplicando las siguientes reglas:

1. El primer símbolo para las vocales A, E, I, O y U, será su correspondiente minúsculas (a, e, i, o, u).
2. La N representará el primer símbolo de la vocal Ě.
3. Si la realización es sencilla no se empleará un segundo símbolo.
4. Si la realización es cortada se empleará como segundo símbolo la letra C.
5. Si la realización es quebrada se empleará como segundo símbolo la letra Q.

La transcripción completa de las 18 vocales se muestra en la figura 4.4 que se presenta a continuación.

Vocales	Realizaciones																	
	Sencillas						Cortadas						Quebradas					
	A	E	I	O	U	Ě	A'	E'	I'	O'	U'	Ě'	A'A	E'E	II	O'O	U'U	Ě'Ě
Transcripción	a	e	i	o	u	N	aC	eC	iC	oC	uC	NC	aQ	eQ	iQ	oQ	uQ	NQ

Figura 4.4. Transcripción de vocales del zapoteco de Choapan.

Transcripción Consonantes.

Las consonantes del zapoteco podemos dividir las en dos tipos: las consonantes que se asemejan a las del español y que se transcriben utilizando reglas del SAMPA español, y aquellas que son desconocidas para el español cuya transcripción se realiza asignando un símbolo que no se encuentre en el alfabeto SAMPA español. La transcripción completa para todas las consonantes se muestra en la figura 4.5 y las reglas que se utilizaron para obtener ésta son las siguientes:

1. Para las consonantes B, D y G cambia su transcripción dependiendo de la posición en que se encuentre de tal forma que si la consonante está en el inicio de la palabra se utilizará el símbolo de estas letras en minúsculas (b, d,

- g), pero si se encuentran en medio su representación será en mayúsculas (B, D, G).
- La transcripción de la consonante G sólo se realizará como en el paso uno cuando forme cualquiera de las siguientes sílabas: GA, GUE, GUI, GO o GU.
 - Para el caso de la X el símbolo asociado dependerá de su posición dentro de la palabra; si se encuentra en el inicio se utiliza Ks y si está en medio se utiliza Gs.
 - Las consonantes DZ, DY y TZ se representarán mediante la misma combinación de letras en minúsculas (dz, dy, tz).
 - La consonante Ž se representará con la letra Z en mayúscula.
 - Aunque la letra “ñ” no pertenece al alfabeto del zapoteco de Choapan se agregará su transcripción ya que el sonido asociado aparece en varias palabras y su transcripción será J.
 - Para el resto de las consonantes se emplea la misma transcripción que si se realizara una transcripción SAMPA en español.

		Consonantes																							
		B	C	QU	CH	D	DZ	DY	G	HU	J	K	L	M	N	P	R	S	T	TZ	X	Y	Z	Ž	Ñ
Transcripción		b	k	k	tS	d	dz	dy	g	w	j	k	l	m	n	p	r	s	t	tz	Ks	jj	T	Z	J
		B			D	D		G	G								rr			Gs					

Figura 4.5. Transcripción de las consonantes del alfabeto que pertenece a la variante de Choapan.

Transcripción de Diptongos.

Los diptongos se presentan cuando existen dos vocales juntas así que la transcripción de los diptongos se realizó de forma similar al español, utilizando las siguientes reglas:

- Si el diptongo no contiene las letras I o U en su realización sencilla entonces su representación consistirá en la unión de la transcripción de cada vocal conservando su posición dentro del diptongo.
- Si el diptongo incluye la letra I o U en su realización sencilla éstas cambiarán por los símbolo “j” o “w” respectivamente, mientras que la otra letra utiliza la transcripción de las vocales.

En la tabla 4.1 se muestra en forma general como se realizó la transcripción de diptongos donde: el término <letra> se emplea para representar cualquier vocal diferente de I o U sencilla y la transcripción de esta vocal se representa por el término <transcripción vocal> la cual se extrae de la figura 4.4. Algunos ejemplos de la utilización de las reglas de transcripción de diptongos se encuentran en la tabla 4.2

Tabla 4.1. Reglas de transcripción de diptongos.

Caso	Transcripción
<letra> I	<Transcripción de la vocal> j
I <letra>	j <Transcripción de la vocal>
<letra> U	<Transcripción de la vocal> w
U <letra>	w <Transcripción de la vocal>
<letra> ₁ <letra> ₂	<Transcripción de la vocal> ₁ <Transcripción de la vocal> ₂

Tabla.4.2. Ejemplos de la transcripción de diptongos.

Caso	Transcripción
AU	aw
ËU	Nw
IA	ja
EA'	eaC
IU	jw

Utilización de la Transcripción Preliminar.

Antes de iniciar la segmentación se necesita crear un diccionario que contenga como mínimo todas las palabras en zapoteco que aparecen en los audios con su respectiva Transcripción Preliminar a un lado. El diccionario debe cumplir con las siguientes reglas:

1. Las palabras se encuentran en mayúsculas y ordenadas alfabéticamente.
2. Se pueden emplear las letras Ñ o Ë en las palabras.
3. No repetir palabras.
4. Escribir la palabra en mayúsculas seguida de uno o más espacios y colocar la transcripción de cada letra de la palabra separada por un espacio.

Un ejemplo de la forma en cómo se realiza la transcripción para el diccionario se presenta en el Anexo E, el cual contiene la transcripción para cada una de las palabras que aparecen a la base de datos de la variante de Choapan del estado de Veracruz.

Segmentación Automática HTK.

HTK (Hidden Markov Model Toolkit) permite realizar la segmentación automática a partir de utilizar “Modelos Ocultos de Markov”. Este proceso consiste en dos etapas, la primera se encarga del entrenamiento de los modelos de Markov (definiendo y configurando uno por cada fonema), el resultado obtenido es introducido en una segunda etapa para así obtener la segmentación. En la figura 4.6 se observa el funcionamiento de la segmentación automática HTK donde la señal de audio y el diccionario de la transcripción son introducidos al software, éste realiza una etapa de entrenamiento y segmentación para el cálculo de los límites de cada fonema (dónde inicia y termina cada uno de ellos), a fin de que se puedan extraer los audios por medio de algún programa como Praat [38].

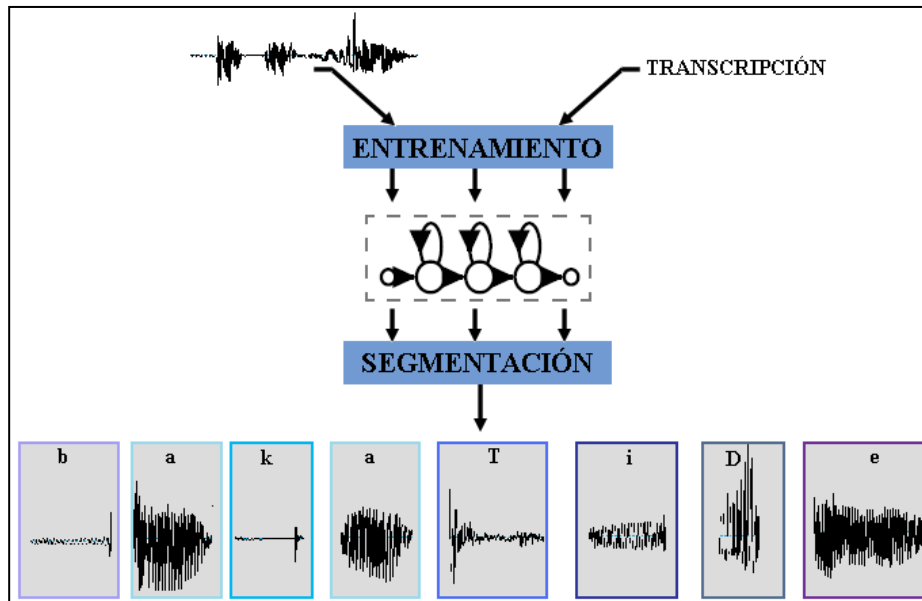


Fig.4.6. Funcionamiento del software HTK para realizar la segmentación de los audios.

La información detallada de cada uno de los pasos para realizar la segmentación HTK se encuentra en el Anexo F, a continuación sólo se presenta un resumen del proceso de segmentación HTK [39]:

- a) Verificar que los audios de la base de datos están en formato .wav, además que el diccionario de transcripción se encuentre ordenado alfabéticamente y con un tamaño mínimo equivalente a la cantidad de palabras que aparecen en los audios.
- b) Crear los archivos de transcripciones; se genera un archivo de texto (extensión .txt) que contenga en forma escrita el contenido de los audios de la base de datos y también se construye un archivo por cada uno de los audios con su información (extensión .lab).
- c) Crear las carpetas para administrar los tipos de archivos de la segmentación HTK. En total son 4: **dbase** que contiene la base de datos, **scr** que tiene los scripts o archivos que utilizan los comandos, **data** contiene los archivos de transcripción y **models** almacena cada uno de los modelos HMM resultado de las reestimaciones que se le hagan a los datos.
- d) Formar una lista de todas las palabras que aparecen en los audios y almacenarlas en el archivo word.list
- e) Crear un diccionario fonético de las palabras de la lista word.list utilizando la herramienta HDMan la cual también entrega la lista de fonemas utilizados.
- f) Obtener dos listas de fonemas utilizados: la primera que incluya el símbolo de silencio (sil) y la segunda que agregue el símbolo de la pausa corta (sp) y el del silencio (sil).

- g) Para cada archivo de audio obtener las palabras utilizadas por cada uno de ellos y almacenarlos en un solo archivo denominado word0.mlf (transcripción a nivel palabra).
- h) Realizar una transcripción a nivel fonema de cada una de las palabras de los audios mediante la utilización de la herramienta HLEd y el resultado se almacenarlo en un solo archivo denominado phone0.mlf.
- i) Almacenar en un archivo de texto las direcciones de los audios a segmentar.
- j) Realizar una parametrización de las señales en secuencias de características mediante la utilización de la herramienta HCompV y reproducir este modelo inicial para cada uno de los fonemas que se utilizan sin incluir la pausa corta (sp).
- k) Realizar tres reestimaciones utilizando la herramienta HERest.
- l) A partir de la última reestimación agregar el modelo HMM correspondiente al símbolo de la pausa corta (sp) y enlazarlo con la herramienta HHed.
- m) Incluir la pausa corta (sp) en el archivo word0.mlf.
- n) Realizar un par de reestimaciones.
- o) Agregar al diccionario creado por HTK la línea de texto “sil sil”, para posteriormente realizar una realineación de los datos mediante la herramienta HVite con lo que termina la etapa de entrenamiento.
- p) Realizar un par de reestimaciones.
- q) Utilizar la herramienta HVite para terminar con la etapa de segmentación.
- r) Aplicar los scripts para crear los archivos que contienen los límites de las palabras y fonemas segmentados.

Para verificar los resultados de la segmentación automática HTK se utilizó el programa Praat que permite abrir el archivo de audio junto con su segmentación y así poder observar los límites de cada fonema de tal forma que si existe un error en éstos podemos modificarlos como se muestra en la figura 4.7.

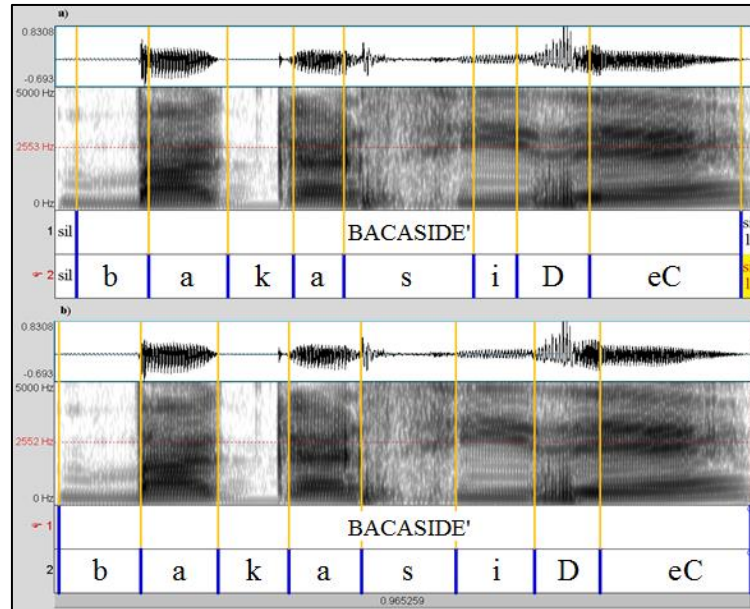


Fig.4.7. Segmentación de la palabra BACASIDE' (muchos): en (a) se muestra la palabra segmentada con grandes deficiencias en los límites de cada fonema. En (b) se muestra la misma palabra con la redefinición de los límites mejorando así su segmentación.

Extracción de Audios.

Los archivos TextGrid son el resultado de la segmentación ya que contienen los límites de cada fonema de la base de datos segmentada y por medio de un script que se implementa en Praat se extrae cada fragmento de audio almacenándolo en su respectiva carpeta creada para cada tipo de fonema.

4.3. Determinación de Fonemas.

La determinación de los fonemas es un paso opcional que se realiza después de la segmentación y se emplea en aquellas vocales que tienen un alto grado de incertidumbre sobre el tipo de realización (cortada, quebrada o sencilla) a la que pertenecen. Este paso se aplica a sonidos que provienen de la base de datos de la variante de Choapan del estado de Veracruz, ya que al ser realizada a partir de la lista Swadesh el hablante proporcionó una opción de la posible ortografía de la palabra en zapoteco, mas no tenemos una alta certeza de que sea la correcta; por lo tanto se realizó una búsqueda en los diccionarios de la lengua zapoteco de la variante de Choapan y de la variante Zoogocho, con lo que se pudo verificar la ortografía de 146 palabras [10, 40, 41]. Las 42 palabras restantes requieren la determinación del tipo de realización de las vocales que se encuentran dentro de ellas con el fin de incrementar la cantidad de ejemplos que se integren en la etapa de análisis.

El método empleado para determinar fonemas fue DTW (Dynamic Time Warping) [42, 43], el cual es muy empleado en el reconocimiento de voz ya que permite comparar señales de audio mediante una métrica que mide la similitud entre éstas, donde entre mayor sea el resultado menos se asemejan los audios y viceversa, entre menor sea este

valor más similares son ellos. Para mayor información sobre la definición del método DTW se puede consultar el Anexo G.

El método DTW utiliza dos tipos de audios:

- Prototipo: son ejemplos de vocales que fueron extraídos de palabras cuya ortografía se conoce.
- Prueba: corresponde al sonido vocálico extraído de alguna de las cuarenta y dos palabras desconocidas. Los sonidos prueba se agrupan de acuerdo al tipo de vocal (A, E, I, O, U) sin tomar en cuenta el tipo de realización, además no se forma un grupo de fonemas desconocidos para la vocal Ë ya que el hablante no la utiliza en ninguna de las palabras con alta incertidumbre. No se debe olvidar que cada palabra tiene dos repeticiones en la base de datos así que se deben utilizar ambos sonidos como archivos de prueba.

La experimentación con DTW se realizó mediante un programa escrito en Matlab R2012b [44] utilizando las funciones presentadas en [45] para el cálculo de DTW. Este programa abre y analiza los audios de prueba entregando el resultado en una tabla de costos como se muestra en el diagrama de la figura 4.8. Se realizaron cinco experimentos que corresponden a cada uno de las cinco tipos de vocales desconocidas y a continuación explicamos cada uno de ellos:

1. Vocal A. El experimento se realizó en dos partes; la primera donde los audios de prueba corresponden a sonidos de la vocal A sencilla que son comparados contra 15 audios prototipo (cinco por cada realización de la vocal). La segunda parte del experimento utilizó como audios de prueba los que corresponden al diptongo AA y éstos se compararon contra 15 audios prototipo de la primera parte más un par de sonidos extra que ejemplifican este diptongo.
2. Vocal E. Los audios de prueba corresponden a dieciséis sonidos que se comparan contra los audios prototipo que corresponden a sonidos tanto de la vocal e y ë en sus tres diferentes realizaciones (debido a la semejanza acústica entre las vocales), en total son veinte audios.
3. Vocal I. Este experimento se utilizaron como audios de prueba 42 sonidos que se compararon contra 14 audios prototipo (5 de la realización sencilla, 5 de la realización cortada y 4 de la realización quebrada)
4. Vocal O. Este experimento consistió de 4 audios de prueba donde dos corresponden a una realización sencilla y los otros dos al diptongo “OO” que se compararon contra 9 audios prototipo, 5 para la realización sencilla y 4 para la realización quebrada. No existen ejemplos en la base de datos de la realización cortada para esta vocal.

5. Vocal U. El experimento consistió de 10 audios de prueba se compararon contra 10 audios prototipo, 5 para la realización sencilla y 5 para la realización cortada, recordando que no existen ejemplos en la base de datos de la realización quebrada para esta vocal.

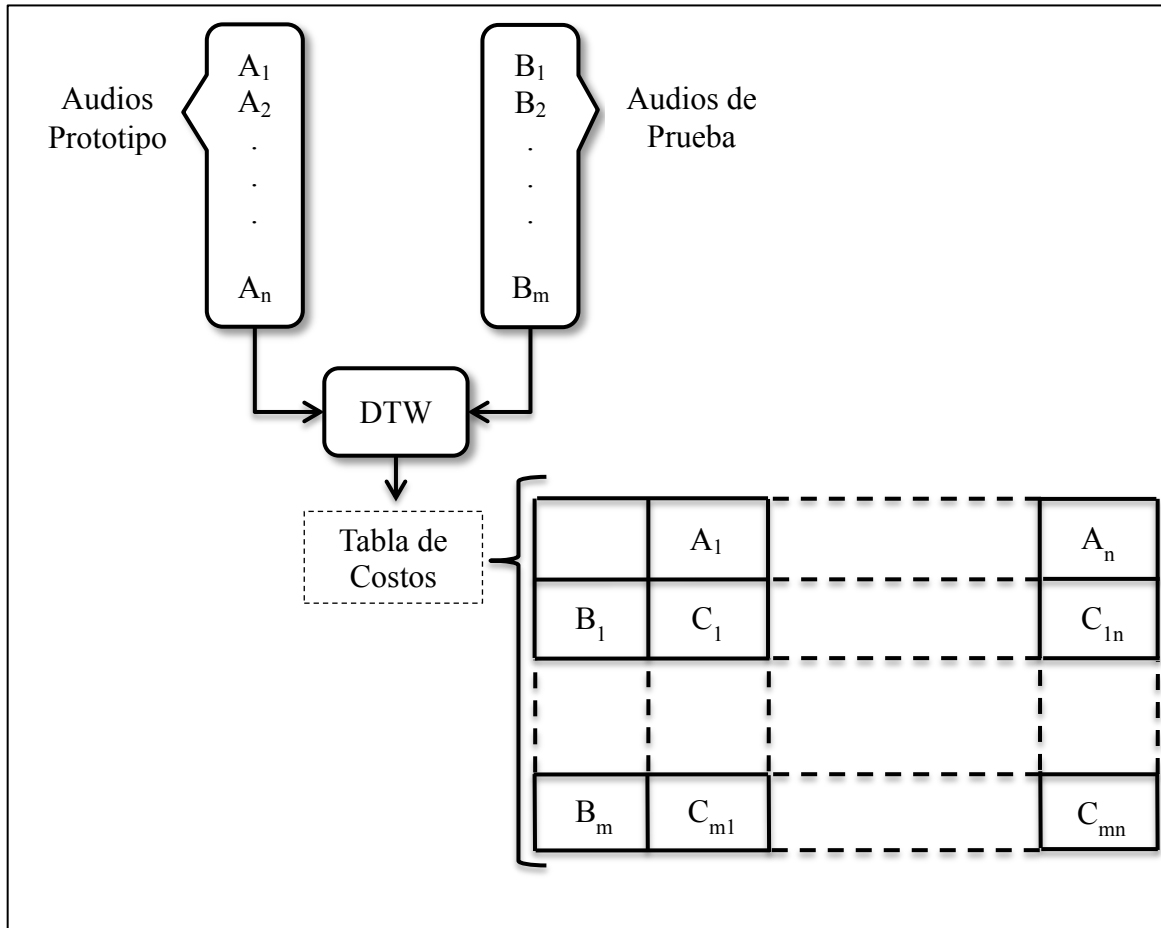


Fig.4.8. Los n audios de prueba se comparan con los m audios prototipo utilizando el método DTW y el resultado se coloca dentro de una tabla de costos.

Para cada una de las pruebas se obtuvo una tabla de costos, la cual contiene los valores de los costos obtenidos a partir de encontrar la trayectoria más corta según DTW para los audios evaluados. En la figura 4.8 se muestra el formato de la tabla, la cual cuenta con las siguientes características:

- “ n ” columnas, una por cada sonido prototipo
- “ m ” renglones, una por cada sonido de prueba
- La tabla cuenta con “ $n \times m$ ” celdas, donde cada de ellas contiene el costo resultado de comparar los sonidos n y m .
- La realización resultante para el sonido de prueba “ n ” corresponderá al tipo de realización del sonido prototipo que tenga el menor costo.

Los sonidos prototipo y prueba se identifican dentro de la tabla con la siguiente sintaxis <TipoVocal><R> <P > <Pl> <Rep>, donde:

- <TipoVocal> es cualquiera de las seis vocales (A, E, I, O, U).
- <R> es la realización de la vocal; C para la cortada, Q para la quebrada y para la sencilla se omite este término.
- <P> es la posición o lugar en que se encuentra este grafema dentro de la palabra, tomando en cuenta que el primer grafema de la palabra siempre tiene la posición dos.
- <Pl> es la palabra en español de la que fue extraído el sonido.
- <Rep> es la repetición de la palabra como fue almacenada en la Base de datos.

Se debe tomar en cuenta al seleccionar el tipo de realización de la vocal dentro de la palabra que cada una de ellas posee dos repeticiones, lo que conlleva a que existan dos sonidos que se representan por el mismo grafema. Por lo tanto para elegir la realización de la vocal se siguen los siguientes pasos:

1. Dentro de la tabla de costos se identifican las dos repeticiones del sonido que pertenecen a la misma palabra.
2. Se observa la realización resultante para cada una de las repeticiones y se toman sus costos.
3. Si esta realización es la misma en ambos sonidos entonces el sonido desconocido se convierte en ese tipo de realización.
4. En otro caso, si la realización es diferente entonces se identifica cuál de las repeticiones tiene menor costo y se toma como un valor base.
5. La repetición con el costo mayor cambia su realización a la del sonido prototipo con el segundo costo más bajo.
6. Se vuelven a comparar las realizaciones
7. Si coinciden éstas en el mismo tipo de realización entonces la vocal adopta ese sonido.
8. En otro caso, si son diferentes las realizaciones se repite el paso 5 pero se invierten los papeles, es decir, se cambia la repetición de menor costo y permanece constante la de mayor valor.
9. Se repiten los pasos seis y siete
10. En otro caso, si son diferentes las realizaciones entonces se varían ambas repeticiones buscando para cada una de ellas la segunda realización con más bajo costo.
11. Se repiten los pasos seis y siete.
12. Si son diferentes entonces consideramos a este sonido como indeterminado (I), ya que el sonido en ambas repeticiones se asemeja a realizaciones diferentes a pesar de que se tomó el segundo audio más semejante.

A continuación se muestran algunos ejemplos de cómo funciona el método para determinar la realización de una vocal.

Ejemplo1. Se tienen los sonidos de la vocal A que se encuentra dentro de la palabra RIÑA' (morder) y que se compara contra seis audios prototipo, dos por cada realización. La tabla 4.3 muestra los costos finales.

Tabla4.3. Tabla de costos para el tipo de vocal A en una realización inicial "cortada" de la palabra RIÑA', la cual permanece constante después del análisis.

SONIDO	aC5cabello1	aC5que1	a3amarillo1	a3arbol1	aQ3partir1	aQ3verde1	COSTO	SONIDO
aC5morder1	6.3539	3.6529	9.5877	7.6554	6.3125	28.4047	3.6529	aC5que1
aC5morder2	6.0623	6.1196	10.246	7.6298	6.584	29.0332	6.0623	aC5cabello1

Los sonidos de la vocal A' de la palabra RIÑA' (morder) en sus dos repeticiones permanecen como una vocal A' (aC), es decir, en la tabla de costos los valores mínimos corresponden en ambas repeticiones a sonidos de vocales cortadas que se encuentran en las palabras BIDA'N (que) y GUITZA' (cabello).

Ejemplo2. Se tienen los sonidos de la vocal A' que se encuentra en la palabra RIYA' (reir) y que fueron comparadas con seis audios prototipo del tipo de vocal A en sus 3 diferentes realizaciones; los valores resultantes de los costos se encuentran en la tabla 4.4.

Tabla4.4. Tabla de costos de la vocal A' que se segmento de la palabra RIYA' (reir), donde el resultado muestra que la vocal permanece con una realización cortada.

SONIDO	aC 5cabello1	aC 5oreja1	a3arbol1	a5tres1	aQ3partir1	aQ3verde1	COSTO	SONIDO
aC5reir1	6.2145	1.6345	6.7981	4.9549	5.2998	25.2808	1.6345	aC 5oreja1
aC 5reir2	3.7729	6.2542	3.1162	4.2497	6.5321	29.6791	3.1162	a3arbol1

Aplicando los pasos anteriores para determinar el tipo de realización se tiene lo siguiente:

1. Comparando las repeticiones se tiene que no coinciden en el tipo realización ya que una es cortada (aC) y la otra es sencilla (a).
2. Dejamos constante la repetición con el costo menor que corresponde a la vocal (aC) de la palabra NAGA' (oreja) que tiene un costo de 1.6345
3. La segunda repetición tiene el costo mayor (3.1162) entonces su realización cambia a otra cuyo costo sea el segundo más pequeño (3.7729) que corresponde a la vocal A' (aC) de la palabra GUITZA' (cabello).
4. Se termina el análisis porque al comparar las realizaciones de ambas repeticiones coinciden con la vocal A'.

Ejemplo3. Se determina la correcta realización para la vocal A que se encuentra dentro de la palabra LUBA' (soplar) que es comparada contra seis audios prototipo, dos por cada tipo de realización obteniendo la tabla de costos 4.5.

Tabla4.5. Tabla de costos de la vocal A' que se segmento de la palabra LUBA' (soplar), donde el resultado muestra que la vocal permanece indeterminada.

SONIDO	aC5ahi1	aC5oreja1	a3arbol1	a5tres1	aQ3beber2	aQ3mano1	COSTO	SONIDO
a5soplar1	4.5513	4.8603	3.761	2.6837	11.2588	2.6553	2.6553	aQ3mano1
a5soplar2	3.1661	2.3025	7.7433	5.759	14.6554	6.3141	2.3025	aC5oreja1

Aplicando los pasos para determinar un tipo de realización se tiene lo siguiente:

1. Comprobando las realizaciones en ambas repeticiones, se tiene que son diferentes.
2. Dejamos constante la segunda repetición de la vocal porque tiene el menor costo (2.3025) que corresponde a la vocal A' (aC) de la palabra NAGA' (oreja)
3. Cambiamos la realización de la primera repetición al sonido que tenga el segundo costo más bajo (2.6553), que corresponde a la vocal A (a) de la palabra TZONAN (tres)
4. Comparando las realizaciones de ambas vocales se tiene que son diferentes, una es cortada y la otra quebrada.
5. Dejamos constante la realización de la primera repetición (2.6553).
6. Cambiamos la realización de la segunda repetición que tiene un costo inicial de 2.3025 al sonido de la realización que tenga el segundo costo más bajo (3.1661).
7. Comparando las realizaciones en ambas repeticiones se tiene que son diferentes
8. Para ambas repeticiones cambiamos la realización al mismo tiempo a la que tenga el segundo costo más bajo; la repetición uno es asignada a la vocal A (a) de la palabra TZONAN (tres) con un costo de 2.6837 y la repetición dos es asignada a la vocal A' (aC) de la palabra LENA' (ahí) con un costo de 3.1661.
9. Comparando las realizaciones de ambas repeticiones se observa que son diferentes, por lo tanto este sonido se considera indeterminado (I).

4.4. Extracción de Características.

La Extracción de Características consiste en obtener para cada uno de los sonidos de las vocales sus características acústicas como son: la duración, la frecuencia fundamental, la intensidad y los valores de las primeras cinco formantes, buscando que éstas permitan identificar cada uno de los tipos de vocales (A, E, Ę, I, O, U) y sus realizaciones (cortada, quebrada, sencilla). El valor promedio de las características se obtiene de forma automática mediante un programa en el software Praat, el cual entrega como resultado un archivo de texto (extensión .txt) que tiene el formato como él se muestra en la figura 4.9. Este archivo puede almacenarse con la extensión arff para poder abrirse en el software Weka [46] o importar su información a otros programas, como por ejemplo las hojas de cálculo.

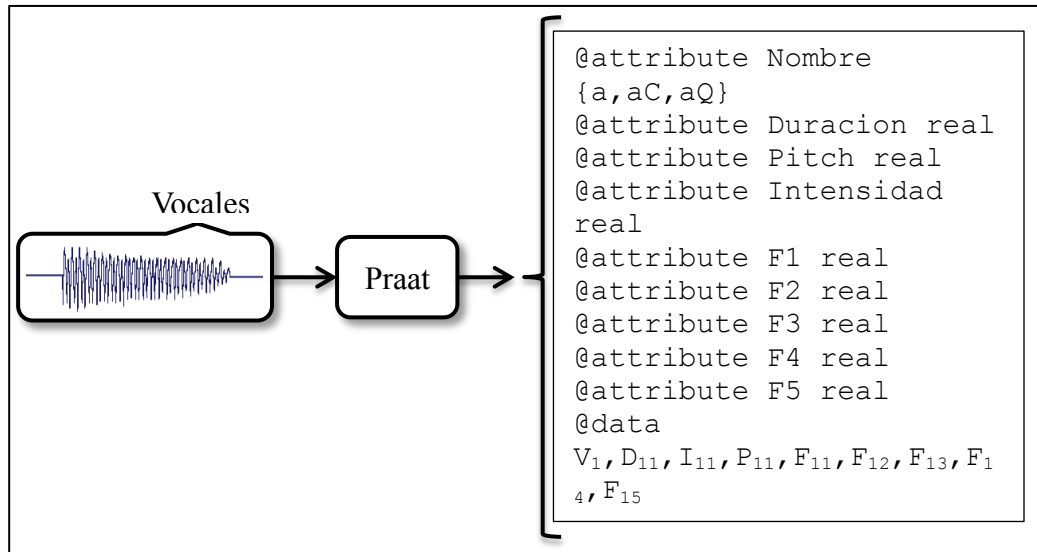


Fig.4.9. Diagrama interno del paso: Extracción de Características.

Sonidos.

Los sonidos a utilizar pertenecen a los sonidos de las 18 vocales de la variante de Choapan obtenidos de la segmentación de las bases de datos para los estados de Veracruz y Oaxaca.

Características a Extraer.

Praat es un software que permite para cada vocal extraer de forma automática el valor promedio de las características: duración, intensidad, pitch y las primeras cinco formantes. La forma en cómo se obtienen éstas se presenta en el anexo H, donde además se indica la instrucción que permite calcularlas dentro del script. [38, 47]

Duración.

Es la cantidad de tiempo que permanece o existe un fonema en el medio, su unidad de medida es el segundo (seg) [48].

Intensidad.

Representa la cantidad de energía de la señal de habla en función del tiempo, su unidad de medida son los decibeles (dB) y está estrechamente relacionada a la amplitud de la señal [48, 11]. Se obtiene a partir de realizar un recorrido (ventaneo) de la señal, calculando su intensidad en cada una de las ventanas y posteriormente el promedio de estas.

Frecuencia Fundamental (F_0).

La F_0 , también es conocida como tono, se define como la frecuencia con que vibran las cuerdas vocales y está relacionada principalmente con señales cuasi periódicas (sonoras) provenientes de sonidos de vocales, su unidad de medida es el Hertz (Hz) [42, 48, 11]. Existen muchas formas de calcular este valor, en el caso del software Praat se utilizo un algoritmo basado en el método de autocorrelaion para obtener la frecuencia

fundamental mediante un ventaneo de la señal del habla [49], para que posteriormente se calcule el promedio con el valor obtenido en cada ventana.

Formantes.

Las formantes se definen como la serie de armónicos producto de las resonancias del tracto vocal al producir un sonido sonoro [42, 48]. En la distribución y estructura de éstos se concentra la mayor parte de la información físico - acústica transportada por la señal del habla que permite la comprensión del mensaje oral contenido en ella. Los cinco primeras formantes son las utilizadas como una forma de identificación de las vocales; la primera (F1) y segunda (F2) están estrechamente relacionadas con la forma del tracto vocal tiene al producir el sonido, la tercera (F3) se relaciona con pocos sonidos y se enfoca especialmente a la calidad del fonema mientras que las formantes de rango más alto (F4, F5) se usan para determinar la calidad de la voz [11, 50]. En el caso de Praat la forma en como se obtienen las formantes es mediante un ventaneo de la señal donde a cada una de estas se aplica el algoritmo de Burg para encontrar los Coeficientes de Predicción Lineal (LPC) [51]. Por ultimo para cada una de las formantes se realiza un promedio con los valores obtenidos en cada ventana.

4.5. Análisis.

El Análisis es el último paso de la metodología donde se busca identificar cada una de las vocales a partir de las características extraídas en el paso anterior, es decir, a partir de los valores del pitch, las formantes, duración e intensidad se busca poder clasificar entre los diferentes tipos de vocales (A, E, Ë, I, O, U) y sus realizaciones (cortada, quebrada, sencilla) mediante algún método de inteligencia artificial. El análisis se divide en dos partes: Representación Gráfica y Métodos de Reconocimiento de Patrones.

Representación Gráfica.

En esta parte del análisis permite identificar los diferentes tipos de vocales y sus realizaciones mediante la obtención de las gráficas de las características que fueron extraídas anteriormente.

Formantes.

En muchas lenguas del mundo se utiliza las primeras dos formantes para observar la forma en cómo se agrupan las vocales mediante la gráfica F1 contra F2 (F1 vs. F2), ya que estas características están asociadas a la articulación que realiza el tracto vocal al producir el sonido [11]. De tal forma se puede realizar la gráfica para determinar la agrupación de los seis tipos de vocales del zapoteco y después para las 18 vocales (tomando en cuenta sus realizaciones) que provienen de ambas bases de datos: de la variante de Choapan del estado de Veracruz y del Distrito de Choapan.

Diagramas de caja.

Se realiza un diagrama de caja para cada una de las características (F₀, intensidad y duración) con el fin de que alguna de éstas sea útil para identificar entre las 18 vocales del zapoteco. Además de que se realizara para los audios de ambas bases de datos (de la variante de Choapan del estado de Veracruz y de la variante del distrito de Choapan).

Métodos de Reconocimiento de Patrones.

El archivo de texto que contiene las características se puede guardar con la extensión .arff y abrirse en el software Weka3-7-8. Este programa permite realizar el análisis implementando diversos métodos de aprendizaje maquina los cuales son utilizados para el proceso de clasificación de vocales y sus realizaciones, además de que se observó el comportamiento de éstos para las vocales E y Ë que tienen una gran semejanza acústica. En el proyecto se usó: perceptrones multicapa (MLP) y máquinas de soporte vectorial (MSV) [52], este último empleo el kernel radial (RBF) y el kernel polinomial (poly); ya que son de los métodos más manejados en el área de análisis y reconocimiento de voz. En el Anexo I se encuentra la descripción de ambos métodos. Tanto para MSV y MLP se utiliza la opción de validación cruzada con un valor de 10.

El análisis debe realizarse tanto para las vocales que provienen de la base de datos de la variante de Choapan del estado de Veracruz y para la que corresponde al distrito de Choapan. Las pruebas que se realizaron sobre los datos se dividen en dos, de acuerdo a la cantidad de clases que se usan en los clasificadores y son:

1. Seis clases que corresponden a los tipos de vocales (A, E, Ë, I, O, U).
2. Dieciocho clases que corresponden a las 18 vocales (A, A', A'A, E, E', E'E, Ë, Ë', Ë'Ë, I, I', I'I, O, O', O'O, U, U', U'U).

Para cada uno de los métodos de aprendizaje maquina se realizaron pruebas de acuerdo a la cantidad de clases a identificar (6 o 18). Estos métodos son:

1. MLP
2. MSV con kernel polinomial
3. MSV con kernel radial

Los tres métodos de aprendizaje maquina varían los atributos (las primeras cinco formantes) que utilizan. De tal forma que para cada método se llevaron a cabo la siguiente serie de experimentos:

1. F_1 y F_2 (F_{1-2})
2. F_1 , F_2 y F_3 (F_{1-3})
3. F_1 , F_2 , F_3 y F_4 (F_{1-4})
4. F_1 , F_2 , F_3 , F_4 y F_5 (F_{1-5})

En la figura 4.10 se observa el diagrama general de cómo se llevan a cabo las pruebas para los sonidos que corresponden a la base de datos de la variante del estado de Veracruz, en total son 24 experimentos que también tienen que realizarse para la base de datos del distrito de Choapan.

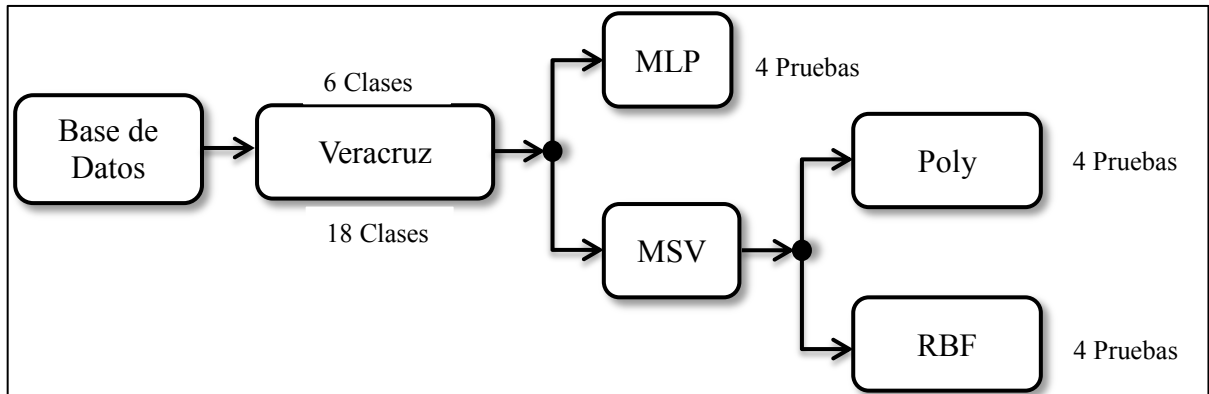


Fig.4.10. Diagrama de pruebas a realizar con los metodos de aprendizaje maquina.

4.6. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).

Es el paso de la metodología que se centra en buscar un símbolo fonético en el alfabeto fonético SAMPA que se use para representar la sexta vocal. Tomando en cuenta que un alfabeto fonético es un conjunto de símbolos que se usan para representar los sonidos existentes en una lengua, se buscó el símbolo fonético correspondiente entre vocales, consonantes, tonos (ascendente, descendente, bajo, alto), acentos, etc.

El Alfabeto Fonético Legible por Ordenador (SAMPA por sus siglas en inglés), es un alfabeto joven al lado de IPA (Alfabeto Fonético Internacional), desarrollado entre los años de 1987 al 1989 por un grupo de fonetistas [53, 54]. Este alfabeto no es más que un conjunto de símbolos los cuales permiten representar un sonido dentro del ordenador, resultado de un mapeo de una parte del alfabeto IPA sobre el código ASCII entre el rango de caracteres 33 al 127. En la actualidad idiomas como el español, alemán, italiano, noruego, danés, húngaro, francés, cantones, sueco, griego, portugués, búlgaro, ruso, etc., cuentan con una representación SAMPA. [32]

IPA fue establecido a partir de la organización fonética internacional formada en 1886 y este alfabeto consiste en una notación estándar para la representación fonética de una lengua, esto permite que una enorme cantidad de lenguas lo utilicen como lo son: el inglés, francés, español, italiano, búlgaro, coreano, farsi, etc. [53]. El alfabeto IPA asigna símbolos a las consonantes y vocales. Este último grupo se representa mediante un Trapecio de vocales donde aparecen cada uno de los símbolos que utiliza IPA para representar a las vocales de una lengua [53, 54]. En la figura 4.11 se tiene el trapecio de vocales, donde:

- Front, Central, Back indica la articulación de la lengua al producir el sonido
- Close, Close-mid, Open-mid, Open indica la abertura de la boca cuando se ejecuta el sonido.

- Cuando dos vocales se realizan con la misma articulación de lengua y abertura de la boca, sólo el redondeo de los labios al producirlas representa alguna diferencia entre ellas. Esto es, si la vocal se encuentra en la posición izquierda dentro del trapecio de vocales entonces no se redondean mientras que si está a la derecha del trapecio necesitará redondear los labios para producir el sonido.

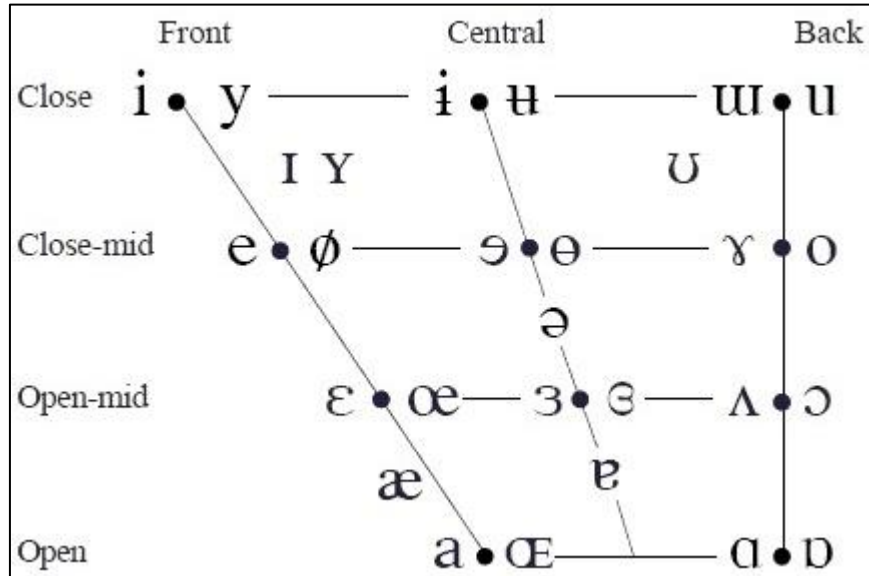


Fig.4.11. Trapecio de Vocales IPA.

SAMPA no posee un mapeo completo de las vocales IPA que aparecen en trapecio, así que con el fin de resolver el problema se realiza una extensión de éste denominada X-SAMPA [55], la cual permite tener la equivalencia para los 28 símbolos de las vocales del trapecio como se muestra en la tabla 4.6.

Tabla.4.6. Tabla de equivalencia de los símbolos de las vocales.

Tabla de Equivalencias	
IPA	SAMPA
ɪ	i
ʏ	y
ɨ	ɪ
ʉ	ɨ
ɯ	ʉ
u	u
ɪ	ɪ
ʏ	ʏ
ɘ	ɘ
e	e
ø	2
ə	@/

ə	8
ɤ	7
o	O
ɛ	E
œ	9
ɜ	3
ɐ	3/
ʌ	V
ɔ	O
æ	{
ɐ	6
a	a
æ	&
ɑ	A
ɒ	Q
ə	@

DTW para encontrar la transcripción fonética de Ę.

La búsqueda de un símbolo fonético que represente a la vocal Ę se realiza mediante la comparación entre los sonidos de dicha vocal contra ejemplos de sonidos que correspondan a los símbolos del trapecio de vocales. La comparación entre las dos señales de audio se realiza mediante el método de reconocimiento de voz DTW (Dynamic Time Warping) y las partes que integran este método se muestra en la figura 4.8

A continuación se explican los tipos de sonido que se utilizan en DTW.

Sonidos Prototipo.

Los sonidos prototipo son un conjunto de audios que contienen ejemplos para cada uno de los símbolos que se encuentran dentro del trapecio de vocales. Estos sonidos son resultado de una búsqueda exhaustiva en diferentes lenguas del mundo, que cuenten con una transcripción fonética IPA o SAMPA. Los pasos para obtener los audios prototipos son:

1. Seleccionar un símbolo fonético IPA o SAMPA.
2. Buscar transcripciones fonéticas IPA o SAMPA en diferentes lenguas o idiomas. [32, 53, 56]
3. Utilizando las anteriores transcripciones se encuentran ejemplos de palabras que contengan el símbolo fonético buscado.
4. Encontrar los audios de la palabra, si esto no es posible descarta y se regresa al paso dos para una nueva palabra. [57]
5. Segmentar manualmente el fonema del símbolo vocálico buscado.

Se seleccionan a lo máximo tres ejemplos por cada símbolo tratando de que exista al menos una voz masculina y femenina entre ellos. También si es posible se eligen aquellos audios que tengan la mejor calidad, utilizando los de baja calidad sólo cuando la cantidad de ejemplos encontrados sea limitada para un símbolo (en el caso de los símbolos “ə” y “e” no se localizaron ejemplos de audios que los ocupen). Además de la calidad de los audios, para seleccionar los sonidos de cada símbolo fonético se utiliza la apreciación acústica buscando los tres audios más similares, de tal forma que en total para todo el conjunto de símbolos fonéticos se tienen 77 elementos que se encuentran especificados en el Anexo J. Cada uno de estos audios se almacenó con la sintaxis SAMPA<Índice><Identificador>, donde:

- SAMPA es un constante que indica que el sonido pertenece a un símbolo fonético.
- <Índice> es un valor alfanumérico proporcionado para identificar el símbolo, los cuales se muestran en la tabla 4.7.
- <Identificador> es un valor alfabético que consiste de tres valores (a, b, c) usados para indicar el número del ejemplo de un símbolo fonético.

Tabla.4.7. Índice asignado a cada símbolo IPA o SAMPA al ser almacenados.

Símbolo		Índice
IPA	SAMPA	
i	i	1
y	y	2
ɨ	1	3
ɥ	}	4
w	M	5
u	u	6
I	I	7
Y	Y	8
ɔ	U	9
e	e	A
ø	2	B
ə	8	C
ɾ	7	D
o	o	E
ɛ	E	F
œ	9	10
ɜ	3	11
ʌ	V	12
ɔ	O	13
æ	{	14
ɐ	6	15

Sonidos de Prueba.

Los audios de prueba pertenecen a ejemplos de sonidos para las vocales Ë y E en sus tres diferentes realizaciones. A continuación se explica cómo se seleccionaron estos audios para ambas bases de datos.

La base de datos de la variante del distrito de Choapan cuenta con la mayor cantidad de ejemplos, por lo tanto se seleccionaron al azar 30 sonidos para las vocales: Ë, Ë', Ë'Ë, E y E'E. En el caso de la vocal E' se seleccionaron 25 de un total de 31 audios.

La base de datos de la variante de Choapan del Estado de Veracruz tiene una cantidad limitada de ejemplos para las vocales Ë y E en sus tres realizaciones, por lo tanto se utilizaron en su mayoría el total de ejemplos que se encuentran en la base de datos con excepción de la vocal E cuyos audios se seleccionaron al azar. La cantidad de sonidos usados para cada una de las vocales son:

- 16 audios para la vocal Ë'
- 8 audios para la vocal Ë'Ë
- 4 audios para la vocal Ë
- 10 audios para la vocal E'
- 6 audios para la vocal E'E
- 30 audios para la vocal E

Herramientas utilizadas para calcular DTW.

Método DTW utilizando la herramienta Matlab.

Las funciones utilizadas del programa Matlab R2012b para el cálculo DTW [45] son:

- `simmx` es la función que calcula la matriz de costos utilizando la métrica distancia de coseno.
- `dp` es la función que encuentra la trayectoria que posee el menor costo dentro de la matriz creada por `simmx`.

Método DTW utilizando la herramienta Praat.

También podemos utilizar el software Praat como un medio para calcular el método DTW ya que este programa cuenta con los scripts implementados internamente para su cálculo. Los pasos para utilizar el método DTW son:

1. Se abren los archivos de audio que se desean comparar.
2. A cada uno de los archivos de audio se calcula los coeficientes cepstrales en escala de mel (MFCC); los valores por default que se eligieron fueron 12 coeficientes cepstrales, con un tamaño de ventana de 0.015 segundos y avances de 0.005 segundos.
3. Los dos cepstrales anteriores se seleccionan y se presiona el botón DTW con lo que aparece la ventana denominada "CC: to DTW" donde: se indica que la distancia entre los coeficientes cepstrales sea por default, los puntos de inicio como final coincidan y selecciona la segunda restricción. En la figura 4.12 se muestra la interfaz de Praat para el cálculo de DTW con las opciones antes mencionadas.

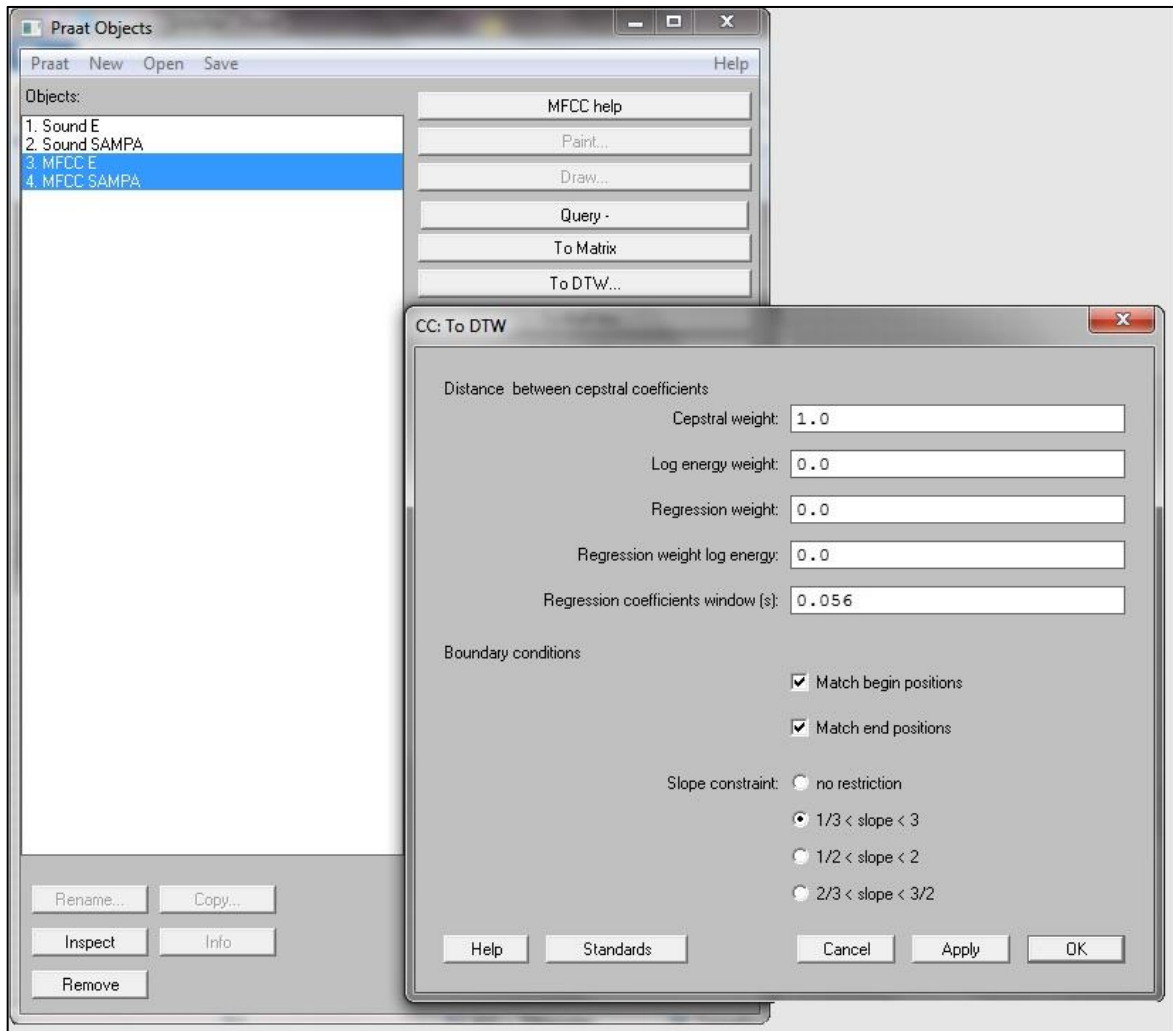


Fig.4.12. Interfaz de Praat para el cálculo DTW, el cual utiliza dos MFCC que pertenecen a los audios E y SAMPÁ.

Experimentación.

El software utilizado para calcular el método DTW es: Matlab y Praat, por lo tanto en general se realizaron los mismos experimentos para ambos programas con las debidas restricciones y características de estos, utilizando audios de ambas bases de datos. Además se comparan ejemplos de la vocal E con el fin de observar si el método DTW puede encontrar diferencias entre ambas vocales. Se llevaron a cabo un total de 24 experimentos.

5. Resultados.

Los resultados obtenidos durante la realización de la metodología que se presentan a continuación están distribuidos en las siguientes secciones: Segmentación, Fonemas, Reconocedor DTW, Determinación de Fonemas, Extracción de características, Análisis y Determinación del símbolo SAMPA de la vocal Ë.

5.1 Segmentación.

La segmentación permite extraer y agrupar cada uno de los fonemas según la vocal a la que pertenezca. Para el caso de la base de datos de la variante del distrito de Choapan el número de fonemas para cada una de las vocales se muestran en la tabla 5.1, la cual mediante tres columnas indica: la vocal, el número de ejemplos de ésta en la base de datos y cuántas de ellas tienen una duración mayor a 60ms. La condición de duración se eligió así por dos razones:

1. Praat requiere sonidos tengan una duración mínima de 60ms para evitar errores al momento de extraer las características
2. Por lo regular las vocales son sonidos que tienen una mayor duración.

Cualquier vocal que no cumpla con la condición de duración se puede considerar como un error de segmentación. De tal forma, y como se muestra en la tabla 5.2, sólo 5790 audios cumplen esta condición, esto significa a que solo el 72.07% de los sonidos fue segmentado correctamente.

Tabla.5.1. Numero de fonemas para cada vocal de la Base de Datos de la variante del distrito de Chopan.

Vocales	Total de Vocales	Duración > 0.06
A	2174	1466
A'	518	460
A'A	64	60
E	1214	735
E'	35	31
E'E	78	67
Ë	356	322
Ë'	730	667
Ë'Ë	143	141
I	1136	691
I'	170	151
I'I	78	67
O	347	225
O'	86	59
O'O	7	7
U	849	598
U'	1	1

U'U	48	42
-----	----	----

Tabla.5.2. Número total de vocales.

	Total Fonemas Vocálicos	Duración > 60ms
Fonemas	8034	5790

5.2. Fonemas.

La base de datos de la variante de Choapan del estado de Veracruz tiene 485 sonidos y la del distrito de Choapan cuenta con 5790 sonidos, que pertenecen a cualquiera de las 18 vocales del zapoteco como se encuentra en la tabla 5.3 la cual presenta la distribución de los audios de acuerdo a la vocal realizada, el porcentaje de éstas con respecto al total audios y el porcentaje con respecto a cada tipo de vocal. Estos sonidos también se pueden agrupar de otras dos formas: de acuerdo al tipo de vocal (A, E, Ë, I, O, U) o al tipo de realización (cortada, quebrada, sencilla), como se muestra en las tablas 5.4 y 5.5 respectivamente.

Tabla.5.3. La columna total indica la cantidad de audios para cada vocal. La columna A indica el porcentaje de la vocal con respecto al total de sonidos utilizados y la columna B señala el porcentaje de la vocal con respecto al tipo de vocal a la cual que pertenece (A, E, Ë, I, O, U).

Vocal	Distrito de Choapan			Playa Vicente, Ver.		
	Total	A [%]	B [%]	Total	A [%]	B [%]
A	1466	25.32	73.28	161	32.78	79.50
A'	460	7.95	23.16	37	6.80	16.50
A'A	60	1.04	3.02	8	1.65	4.00
E	735	12.70	88.23	53	10.93	81.54
E'	31	0.53	3.72	6	1.24	9.23
E'E	67	1.16	8.04	6	1.24	9.23
Ë	322	5.56	28.49	4	0.82	22.22
Ë'	667	11.52	59.03	12	2.47	66.67
Ë'Ë	141	2.43	12.48	2	0.41	11.11
I	691	11.93	76.02	70	14.84	73.47
I'	151	2.61	16.61	24	4.54	22.45
I'I	67	1.16	7.37	4	0.82	4.08
O	225	3.88	77.32	22	4.54	84.61
O'	59	1.02	20.27	0	0	0
O'O	7	0.12	2.40	4	0.82	15.38
U	598	10.33	93.29	60	12.37	76.92
U'	1	0.02	0.16	18	3.71	23.01
U'U	42	0.72	6.55	0	0	0

Tabla.5.4. Cantidad de fonemas vocales de acuerdo al tipo de vocal y su porcentaje con respecto al total de sonidos de las vocales.

Tipo de vocal	Fonemas vocálicos			
	Distrito de Choapan, Oax.		Playa Vicente, Ver.	
	Total de Vocales	Porcentaje [%]	Total de Vocales	Porcentaje [%]
A	1986	34.30	206	41.24
E	833	14.40	65	13.40
Ē	1130	19.52	18	3.71
I	909	15.70	98	20.21
O	291	5.03	26	5.36
U	641	11.07	78	16.08

Tabla.5.5. Cantidad de vocales por tipo de realización y su porcentaje con respecto al total de sonidos de las vocales.

Tipo de realización	Cantidad Total			
	Distrito de Choapan		Playa Vicente, Ver.	
	Total de Vocales	Porcentaje [%]	Total de Vocales	Porcentaje [%]
Sencillas	4037	69.72	370	76.30
Cortadas	1369	23.64	97	18.76
Quebradas	384	6.632	24	4.94

5.3. Reconocedor DTW.

Antes de iniciar el paso de la metodología “determinar fonemas” se realizaron dos pruebas extra, con el objetivo de conocer cómo se comporta el método DTW al utilizar audios de una misma base de datos, es decir, qué porcentaje de audios reconoce correctamente este método.

Resultados de la Prueba I. Se tomaron 100 audios que pertenecen a la base de datos de la variante de Choapan del estado de Veracruz que incluyen ejemplos de todas las vocales en su realización sencilla. La mitad de los audios son archivos de prueba y el resto son prototipo, donde cada vocal tiene la misma representación en ambos tipos de archivo.

En la tabla 5.6 se encuentra el resultado de la prueba I donde se indica que para los 50 audios de prueba 38 fueron correctamente identificados, lo que equivale a que el método DTW reconoció correctamente el 76% de los sonidos.

Tabla.5.6. Resultado del método DTW para la Prueba I.

	Cantidad de Audios
Correctos	38
Incorrectos	12
Total	50

Resultados de la Prueba II. Se tomaron 120 audios que pertenecen a la base de datos de la variante del distrito de Choapan que incluyen ejemplos de todas las vocales en su realización sencilla. La mitad de los audios son archivos de prueba y el resto son prototipo, donde cada vocal tiene la misma representación en ambos tipos de archivo.

En la tabla 5.7 se encuentra el resultado de la prueba II donde se indica que para los 60 audios de prueba 49 fueron correctamente identificados, lo que equivale a que el método DTW reconoció correctamente el 81.67% de sonidos.

Tabla.5.7. Resultado del método DTW para la Prueba II.

	Cantidad de Audios
Correctos	49
Incorrectos	11
Total	60

5.4. Determinación de Fonemas.

Los resultados de esta sección corresponden al paso de la metodología: “Determinación de Fonemas” que consiste en emplear el método DTW, para encontrar todos los fonemas de vocales que son indeterminados y que pertenecen a la base de datos de Choapan del estado de Veracruz. El número de fonemas de las vocales que se encuentran desconocidos o indeterminados para cada palabra se encuentran especificados en la tabla 5.8.

Tabla.5.8. Número de fonemas indeterminados en la Base de datos de Choapan del estado de Veracruz.

Vocales	Cantidad de Fonemas
A	25
E	8
I	21
O	3
U	13

La tabla 5.9 fue creada después de realizar el análisis para determinar fonemas y consiste de aquellas palabras que no fue posible verificar la ortografía, es decir, que desconocemos las letras (vocales y consonantes) que componen a éstas. La tabla se integra por la palabra en zapoteco antes (palabra inicial) y después (palabra final) de determinar las vocales que tenían un alto grado de incertidumbre. Para una mejor visualización de los cambios dentro de las palabras, las vocales analizadas se encuentran

resaltadas en negritas y aquellas que continuaron siendo indeterminadas después del análisis se señalan con el símbolo <I>.

Tabla.5.9. Fonemas finales para las cinco tipos de vocales indeterminados resultado del análisis para determinar fonemas.

No.	Palabra Inicial		Palabra Final
	Español	Zapoteco	
1	Algunos	Nulan na'	Nula'an na'
2	Apretar	Quichan	Qui'cha'an
3	Apuñalar	Quiba	Quiba'
4	Arañar	Iniza	I'ni'za'
5	Atar	Schogan	Schoga'an
6	Calido	Szla	Slza
7	Cazar	Raguaba	Raguaba
8	Golpear	Izana	I'zana'a
9	Gordo	Mallí	Mallí'
10	Jalar	Zan	Za'n
11	Lavar	Guiva	Guiv<I>
12	Morder	Riña'	Ri'ña'
13	Niebla	Lanza	Lanza
14	Nube	Chula	Chula'
15	Reír	Riya'	Ri'ya'
16	Rodilla	Rarriba	Rarri'b<I>
17	Soplar	Luba	Lub<I>
18	Vomitir	Gueba	Guë'ëb<I>
19	Volar	Saa	Sa'
20	Cuerno	Luzubaa	Luzub<I>
21	Achrse	Cuege	Cuege'
22	Caer	Guezo	Guë'z<I>
23	Cuello	Gueru	Gueru'
24	Lago	Elgun	Ë'lgun
25	Nosotros	Neturu	Në'ëtur<I>
26	Respirar	Reziabe	Rë'ëziabe'
27	Blanco	Chiguichi	Chiguichi'
28	Higado	Guirri	Guirri
29	Hincharse	Gui	Gui'
30	Liso	Lili	Lili'
31	Montaña	Guizidau	Gui'zi'dau
32	Sucio	Guchi	Guchi
33	Arena	Roo	Ro
34	Chupar	Zupia	Zupia
35	Oler	Nupian	N<I>pian
36	Pierna	Zun	Z<I>n

La tabla 5.10 indica la cantidad de archivos de audio por cada vocal antes y después del análisis realizado para “Determinar Fonemas”. Hay que tener en cuenta que la base de datos posee dos repeticiones por cada palabra grabada lo que conlleva a que cada fonema que la integra cuenta con dos sonidos y por lo tanto el número de éstos se duplicó, con excepción de las palabras: ŽE (grande) y LAGA’ (ancho) que sólo cuentan con una repetición, así que la cantidad de audios de las vocales A, A’ y E se verán afectadas en menos un valor.

Tabla.5.10. Número de vocales extraídas de las grabaciones de la base de datos de la variante de Choapan del estado de Veracruz antes y después de realizar el análisis. La cantidad de las vocales A, A’ y E se encuentran señaladas con un *.

Tipo de Realización	Cantidad de vocales extraídas de las grabaciones			
	Una repetición			Dos repeticiones
	Fonemas Analizados (desconocidos)	Fonemas Conocidos	Fonemas Analizados mas Conocidos	Total
A	8	80	88	175*
A’	8	17	25	49*
A’A	4	4	8	16
E	1	27	28	55*
E’	2	3	5	10
E’E	0	3	3	6
Ě	0	2	2	4
Ě’	2	6	8	16
Ě’Ě	3	1	4	8
I	8	36	44	88
I’	13	11	24	48
I’I	0	2	2	4
O	2	11	13	26
O’	0	0	0	0
O’O	0	2	2	4
U	9	30	39	78
U’	1	9	10	20
U’U	0	0	0	0

Los resultados se muestran de forma detallada en el Anexo K para cada una de las pruebas realizadas al "Determinar Fonemas".

5.5. Extracción de características.

Las características extraídas para cada uno de los sonidos de las vocales fueron: duración, pitch, intensidad y las cinco primeras formantes. En las siguientes tablas 5.11 y 5.12 se muestra el valor promedio y su desviación estándar por característica (aunque sólo se encuentran los valores de las dos primeras formantes) de cada una de las 18 vocales de ambas bases de datos.

Tabla.5.11. Valores promedio (X) y desviación estándar (DE_x) de las características extraídas para la variante de Choapan del estado de Veracruz de un total de 607 audios.

	Duración [s]		Pitch [Hz]		Intensidad [dB]		F1 [Hz]		F2 [Hz]	
	X	DE _x	X	DE _x	X	DE _x	X	DE _x	X	DE _x
A	0.14	0.05	201.41	23.25	69.76	5.81	819.90	67.81	1786.10	137.90
A'	0.15	0.05	199.05	26.75	73.97	3.86	859.80	63.24	1821.26	112.78
A'A	0.17	0.13	193.58	41.24	70.33	21.18	873.74	117.80	1729.66	323.46
E	0.14	0.05	203.86	19.74	71.43	20.09	503.68	40.26	2229.04	160.62
E'	0.16	0.04	184.59	49.71	70.31	5.33	513.93	65.23	2400.94	100.25
E'E	0.18	0.07	204.69	8.40	69.79	1.55	587.92	43.22	2336.10	100.15
È	0.12	0.03	223.68	9.64	76.68	2.18	532.96	17.65	2235.36	102.66
È'	0.13	0.05	210.28	12.22	74.94	2.85	551.46	71.32	2187.06	249.62
È'È	0.10	0.03	197.86	7.39	74.37	30.52	546.65	62.33	2265.60	103.37
I	0.13	0.04	218.55	21.18	68.45	3.78	360.69	50.98	2535.82	295.20
I'	0.12	0.05	211.36	34.24	66.90	11.27	380.39	46.69	2540.75	285.80
I'I	0.33	0.07	223.05	8.74	62.12	4.98	441.20	70.88	2500.04	148.90
O	0.14	0.04	204.05	26.62	74.67	3.80	546.25	44.31	1158.16	96.32
O'	-	-	-	-	-	-	-	-	-	-
O'O	0.34	0.07	207.77	11.89	75.32	5.22	584.46	22.69	1102.29	31.76
U	0.13	0.05	224.86	24.31	73.98	20.28	428.45	38.57	1063.10	144.12
U'	0.14	0.05	230.95	22.24	73.87	4.23	460.65	57.01	1082.98	179.36
U'U	-	-	-	-	-	-	-	-	-	-

Tabla.5.12. Valores promedio (X) y desviación estándar (DE_x) de las características extraídas para la variante del Distrito de Choapan de un total de 5970 audios.

	Duración [s]		Pitch [Hz]		Intensidad [dB]		F1 [Hz]		F2 [Hz]	
	X	DE _x	X	DE _x	X	DE _x	X	DE _x	X	DE _x
A	0.119	0.06	137.462	14.91	82.611	4.00	426.344	144.71	1801.885	211.03
A'	0.142	0.06	128.886	40.52	71.075	7.26	408.490	97.75	2004.838	163.53
A'A	0.161	0.07	128.105	40.89	79.436	4.39	440.723	135.94	1892.893	141.81
E	0.121	0.12	132.895	51.15	79.912	8.63	526.153	104.57	1519.927	127.48
E'	0.139	0.12	125.143	51.15	74.933	8.63	554.250	104.57	1456.922	127.48
E'E	0.161	0.11	124.663	42.07	77.643	3.98	641.473	100.64	1509.653	136.12
Ē	0.115	0.05	138.258	13.51	81.635	2.01	381.626	43.54	1980.531	111.30
Ē'	0.194	0.05	122.179	40.96	72.449	8.65	415.173	104.25	2087.531	125.50
Ē'Ē	0.211	0.09	111.392	34.93	76.466	4.14	409.989	62.11	2128.281	81.80
I	0.132	0.07	135.176	18.73	79.098	6.31	299.105	92.15	2221.781	147.26
I'	0.161	0.08	122.349	39.62	67.887	12.14	351.768	156.19	2186.486	116.34
I'I	0.109	0.03	116.385	59.34	78.075	4.04	276.126	16.57	2199.074	73.22
O	0.158	0.10	126.210	13.39	74.613	9.36	486.717	206.91	1289.012	381.19
O'	0.223	0.15	119.714	49.21	56.338	18.89	569.493	235.72	1370.912	358.68
O'O	0.221	0.19	107.087	73.33	73.090	7.11	488.806	52.10	1224.153	101.04
U	0.123	0.05	141.985	16.60	80.380	4.68	330.6112	72.41	1087.017	242.53
U'	0.110	0.00	135.886	0.00	62.465	0.00	450.507	0.00	1107.051	0.00
U'U	0.196	0.10	125.122	22.20	77.758	4.25	345.497	59.92	1091.075	166.23

5.6. Análisis.

Los resultados de esta sección corresponden a los obtenidos en el proceso de análisis de los fonemas de las vocales. Este se divide en dos etapas: la primera consiste en la representación gráfica de las características extraídas y la segunda utiliza estas mismas para realizar una clasificación de las vocales utilizando métodos de reconocimiento de patrones.

Representación Gráfica.

Esta etapa utiliza las características extraídas de cada uno de los audios para tratar de identificar entre los tipos de vocales (A, E, Ē, I, O, U) y sus realizaciones (sencilla, cortada, quebrada). Los primeros resultados de esta etapa consisten de la localización de las dos primeras formantes mediante una gráfica (F1 vs. F2) la cual permite la formación de seis grupos, que corresponden a cada uno de los tipos de vocales (sin tomar en cuenta la realización de la vocal) como se muestran en las figuras 5.1 y 5.2 respectivamente para cada base de datos.

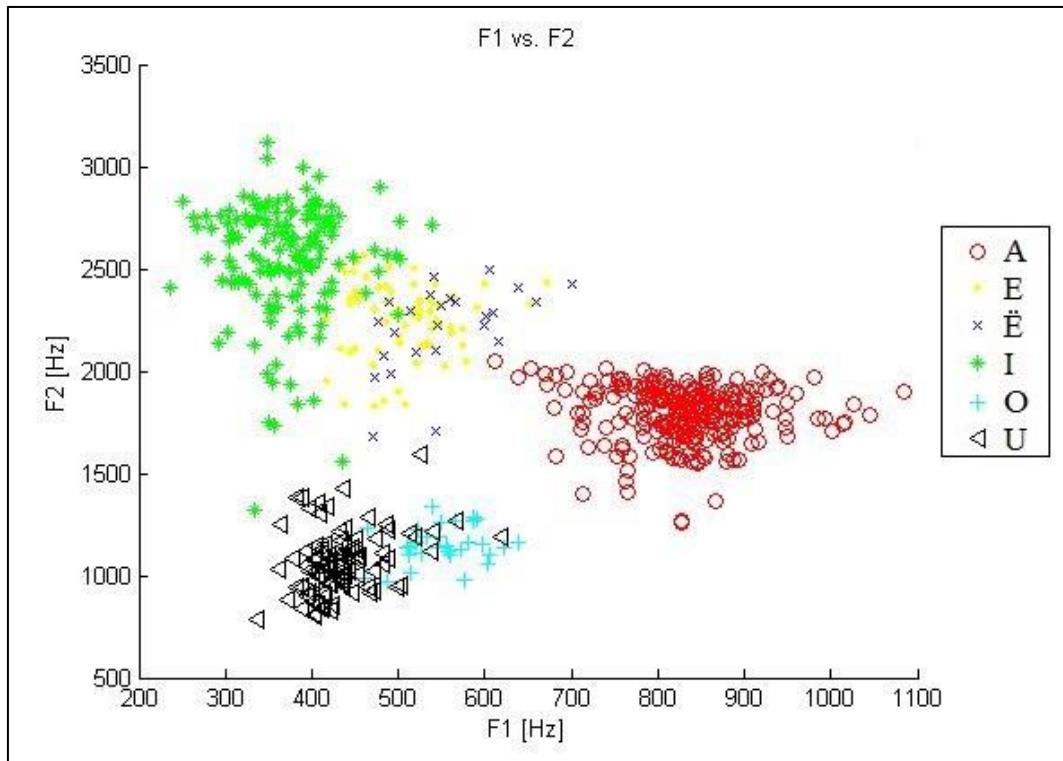


Fig.5.1. Vocales de la variante de Choapan del estado de Veracruz.

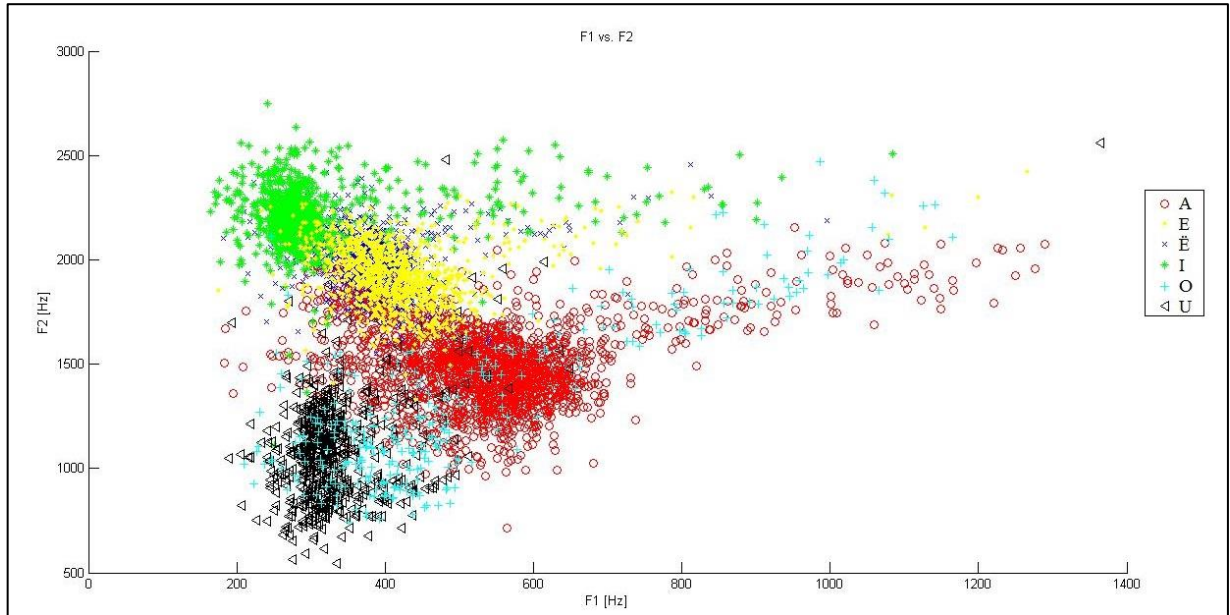


Fig.5.2. Vocales de la la variante del distrito de Choapan.

Las características F1 y F2 permiten agrupar los seis tipos de vocales pero esto no funciona de misma forma para identificar el tipo de realización de cada vocal (cortada, quebrada sencilla). En la figuras 5.3 y 5.4 se presentan la gráfica F1 contra F2 para cada uno de los grupos de las vocales en sus tres tipos de ejecuciones y observamos que no se agrupan las realizaciones, sino todo lo contrario éstas se encuentran distribuidas en la gráfica.

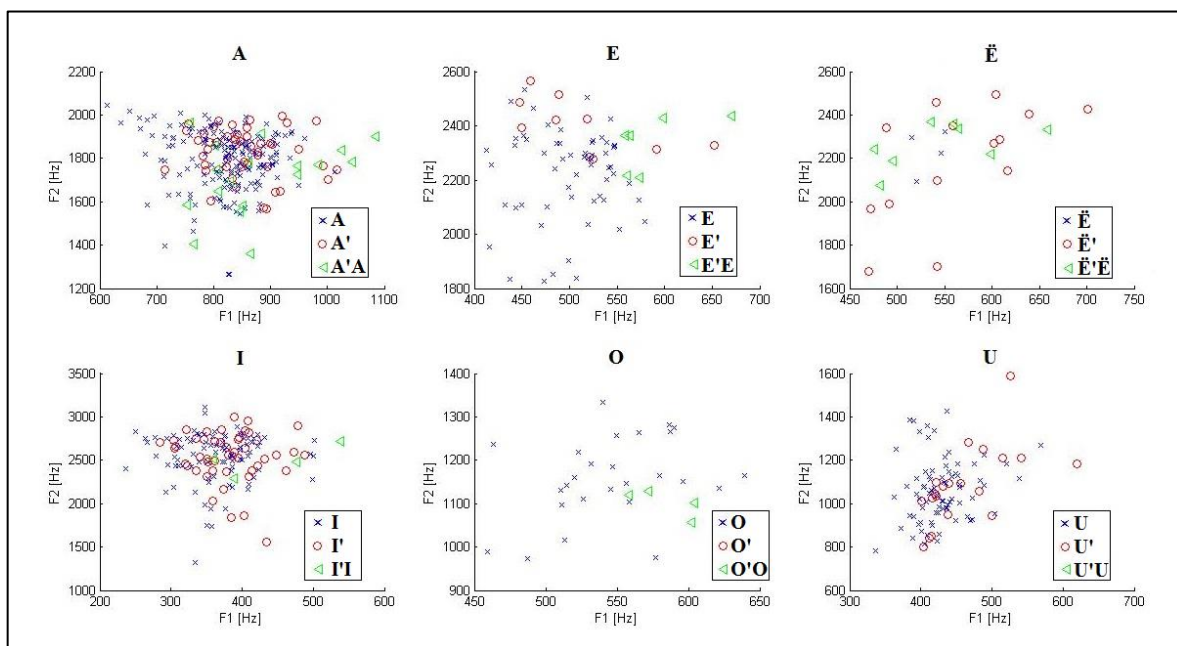


Fig.5.3. Gráfica de las vocales según su realización de la variante de Choapan del estado de Veracruz.

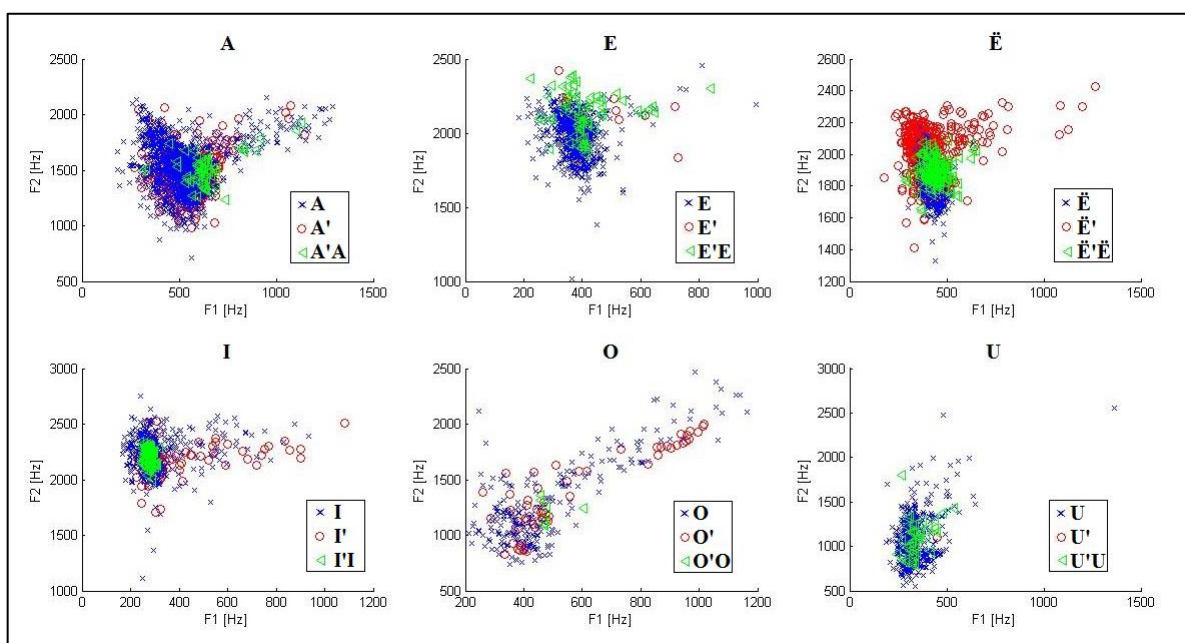


Fig.5.4. Gráfica de las vocales según su realización de la variante del distrito de Choapan.

Las figuras 5.5 y 5.6 muestran la gráfica de caja de las características: frecuencia fundamental, intensidad y duración, que pertenecen a las 18 vocales de ambas bases de datos.

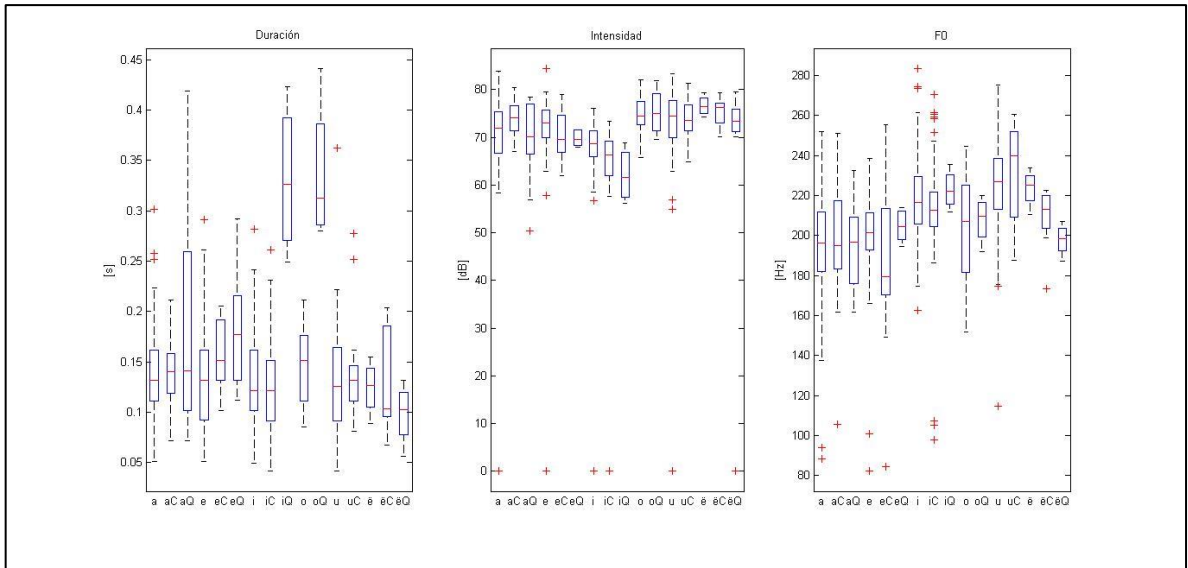


Fig.5.5. Gráficas de caja para la Duración, Intensidad y Frecuencia Fundamental de la variante de Choapan del estado de Veracruz.

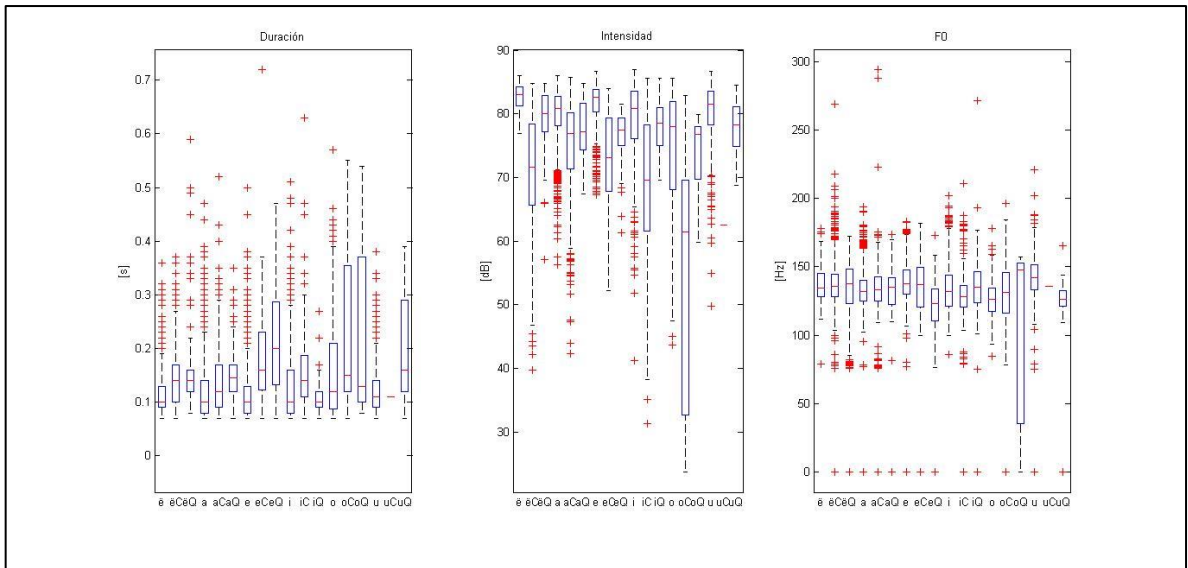


Fig.5.6. Gráficas de caja para la Duración, Intensidad y Frecuencia Fundamental de la variante del distrito de Choapan.

Métodos de Reconocimiento de Patrones.

La otra parte de los resultados corresponde a la clasificación de los datos mediante la utilización de dos métodos de aprendizaje supervisado que son redes neuronales (MLP) y máquinas de soporte vectorial (MSV); este último emplea dos tipos de kernel: polinomial (poly) y base radial (RBF). Todos estos métodos fueron aplicados a los datos de la variante zapoteca de Veracruz y de Oaxaca, utilizando como atributos las cinco primeras formantes (F1, F2, F3, F4, F5) y variando desde la primera hasta la quinta en cada prueba. A continuación se presenta cada una de las series de pruebas donde sólo se muestran los seis mejores resultados.

Vocales de la base de datos de Choapan del estado de Veracruz.

Primera serie de pruebas: Los clasificadores utilizan los seis tipos de vocales como clases, donde el mejor resultado se obtuvo con un MLP que logra clasificar el 91.98% de las instancias correctamente. Utilizando como atributos las primeras cuatro o las primeras dos formantes como se muestra en la tabla 5.13 y en el inciso a de la figura 5.7 se encuentra la matriz de confusión para este resultado.

Segunda serie de pruebas: Los clasificadores utilizan las dieciocho vocales como clases, donde el mejor resultado se obtiene con un MLP que logra clasificar el 67.26% de las instancias correctamente. Utilizando como atributos las primeras dos formantes como se muestra en la tabla 5.14 y en el inciso a de la figura 5.8 se encuentra la matriz de confusión para este resultado.

Tabla.5.13. Seis mejores resultados utilizando los clasificadores MLP, MSV polinomial y MSV radial, para identificar los seis tipos de vocales.

No.	Método	Instancias Correctamente Clasificadas [%]
1	MLP F1-4	91.98
2	MLP F1-2	91.98
3	MLP F1-3	91.32
4	MLP F1-5	90.34
5	MSV Poly F1-4	87.88
6	MSV Poly F1-4	87.88

Tabla.5.14. Seis mejores resultados utilizando los clasificadores MLP, MSV polinomial y MSV radial, para identificar las dieciocho vocales.

No.	Método	Instancias Correctamente Clasificadas [%]
1	MLP F1-2	67.26
2	MLP F1-3	66.77
3	MLP F1-4	66.12
4	MLP F1-5	65.13
5	MSV Poly F1-4	63.33
6	MSV Poly F1-3	63.33

El método que obtuvo el peor resultado en ambas series de pruebas corresponde a un MSV con kernel RBF que utiliza como atributos las primeras dos formantes, ya que sólo logró clasificar correctamente el 50.24% y 28.31% de las instancias respectivamente para cada serie de pruebas. Las matrices de confusión se encuentran en el inciso b de las figuras 5.7 y 5.8 respectivamente para cada uno de estos casos.

a)							b)						
=== Matriz de Confusión ===							=== Matriz de Confusión ===						
a	b	c	d	e	f	<-- clasificado	a	b	c	d	e	f	<-- clasificado
238	0	1	0	1	0	a = a	240	0	0	0	0	0	a = a
0	57	6	8	0	0	b = e	71	0	0	0	0	0	b = e
0	20	8	0	0	0	c = è	28	0	0	0	0	0	c = è
0	3	0	135	0	2	d = i	75	0	0	65	0	0	d = i
0	0	0	0	27	3	e = o	30	0	0	0	0	0	e = o
0	0	0	1	4	93	f = u	98	0	0	0	0	0	f = u

Fig.5.7. Estas matrices de confusión corresponden a la primera serie de pruebas. El inciso a pertenece al método MLP F1-4 y el inciso b al método MSV RBF F1-2.

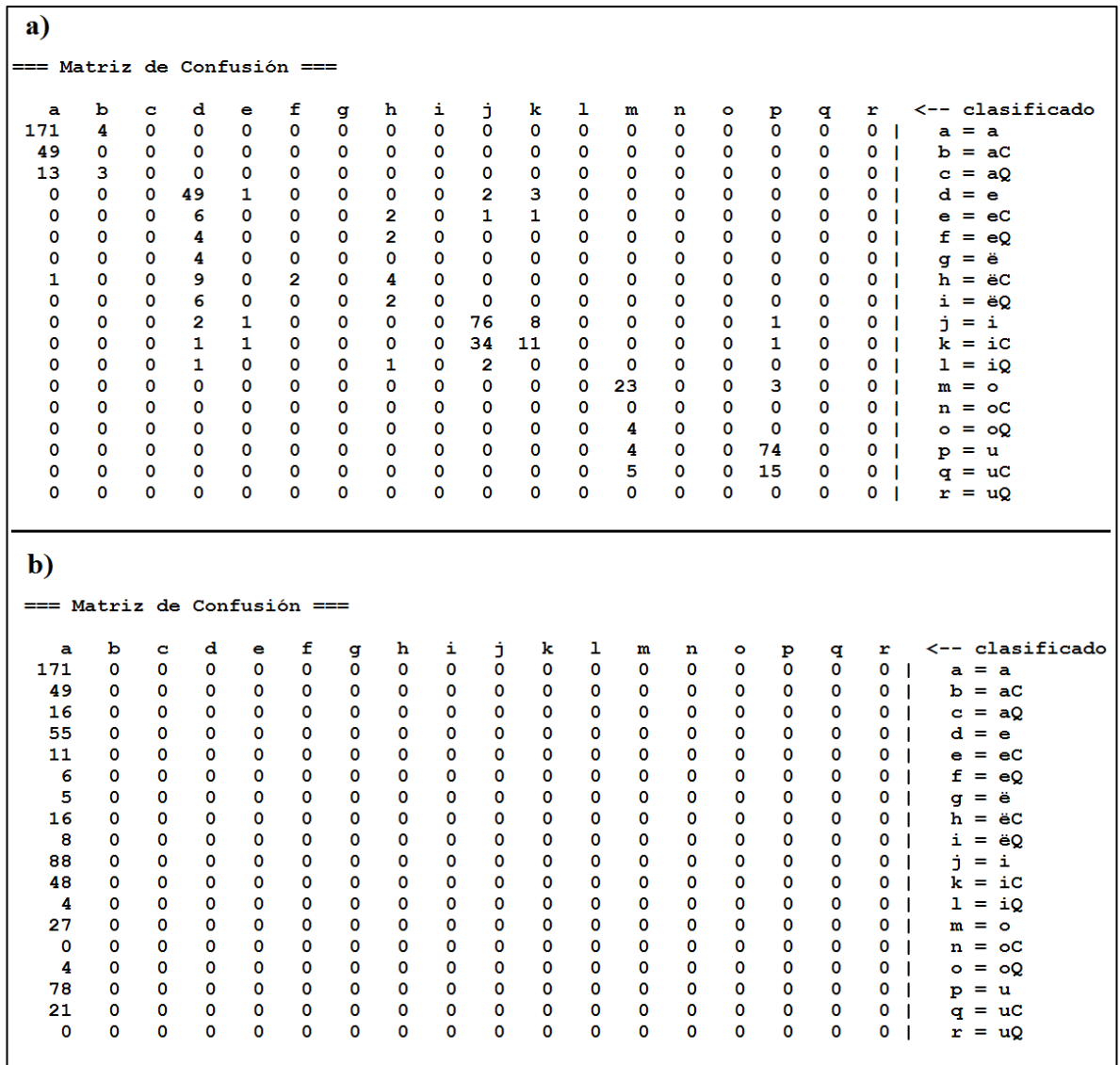


Fig.5.8. Estas matrices de confusión corresponden a la segunda serie de pruebas. El inciso a pertenece al método MLP F1-2 y el inciso b al método MSV RBF F1-2.

Vocales de la base de datos del distrito de Choapan.

Primera serie de pruebas: Los clasificadores utilizan los seis tipos de vocales como clases, donde el mejor resultado se obtiene con un MLP que logra clasificar el 77.36% de las instancias correctamente y la matriz de confusión para este caso se encuentra en el inciso a de la figura 5.9. Utilizando como atributos las primeras cinco formantes como se muestra en la tabla 5.15.

Segunda serie de pruebas: Los clasificadores utilizan las dieciocho vocales como clases, donde el mejor resultado se obtiene con un MLP que logra clasificar el 61.35% de las instancias correctamente y la matriz de confusión para este caso se encuentra en el

inciso b de la figura 5.10. Utilizando como atributos las primeras cinco formantes como se muestra en la tabla 5.16.

El método que obtuvo el peor resultado en ambas series de pruebas corresponde a un MSV con kernel RBF que utiliza como atributos las primeras dos formantes, ya que solo logro clasificar correctamente el 36.94% y 25.32% de las instancias respectivamente para cada serie de pruebas. En el inciso b de las figuras 5.9 y 5.10 se encuentra la matriz de confusión que corresponde a estos casos respectivamente.

Tabla.5.15. Seis mejores resultados utilizando los clasificadores MLP, MSV polinomial y MSV radial, para identificar los seis tipos de vocales.

No.	Método	Instancias Correctamente Clasificadas [%]
1	MLP F1-5	77.36
2	MLP F1-4	77.35
3	MSV Poly F1-4	75.20
4	MLP F1-3	75.17
5	MSV Poly F1-5	74.89
6	SVM Poly F1-3	73.04

Tabla.5.16. Seis mejores resultados utilizando los clasificadores MLP, MSV polinomial y MSV radial, para identificar las dieciocho vocales.

No.	Método	Instancias Correctamente Clasificadas [%]
1	MLP F1-5	61.35
2	MLP F1-4	60.97
3	MLP F1-3	59.64
4	MSV Poly F1-5	59.03
5	MSV Poly F1-4	58.69
6	MSV Poly F1-3	56.53

a)								b)							
=== Matriz de Confusión ===								=== Matriz de Confusión ===							
a	b	c	d	e	f	<-- clasificado		a	b	c	d	e	f	<-- clasificado	
1830	24	79	8	33	12	a = a		1978	0	8	0	0	0	a = a	
24	413	327	67	2	0	b = e		609	0	224	0	0	0	b = e	
55	224	786	62	1	2	c = è		970	0	160	0	0	0	c = è	
0	38	39	829	0	3	d = i		77	0	831	1	0	0	d = i	
101	1	2	5	61	121	e = o		290	0	1	0	0	0	e = o	
14	1	9	3	54	560	f = u		640	0	1	0	0	0	f = u	

Fig.5.9. Estas matrices de confusión corresponden a la primera serie de pruebas para la variante del distrito de Choapan. El inciso a pertenece al método MLP F1-5 y el inciso b al método MSV RBF F1-2.

a)
 === Matriz de Confusión ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	<-- clasificado
1197	161	1	33	0	0	20	13	7	6	2	0	13	9	0	4	0	0	a = a
267	172	0	3	0	0	10	3	0	1	0	0	2	0	0	2	0	0	b = aC
43	12	0	0	0	0	1	0	0	0	0	0	0	4	0	0	0	0	c = aQ
21	0	0	448	0	3	75	132	6	43	2	2	2	0	0	1	0	0	d = e
1	0	0	12	0	0	0	14	0	4	0	0	0	0	0	0	0	0	e = eC
1	0	0	22	0	2	1	26	0	13	2	0	0	0	0	0	0	0	f = eQ
32	4	0	75	0	0	160	34	15	1	0	0	0	0	0	1	0	0	g = ë
6	2	0	237	0	0	45	315	14	34	8	1	4	1	0	0	0	0	h = ëC
8	0	0	33	0	0	31	38	31	0	0	0	0	0	0	0	0	0	i = ëQ
1	0	0	36	0	2	0	35	0	592	10	12	1	0	0	2	0	0	j = i
1	0	0	3	0	1	0	29	0	86	23	8	0	0	0	0	0	0	k = iC
0	0	0	0	0	0	0	3	0	47	12	5	0	0	0	0	0	0	l = iQ
56	12	1	2	0	0	0	0	1	1	1	0	67	5	0	79	0	0	m = o
7	2	1	0	0	0	0	0	0	0	0	0	14	17	0	17	0	1	n = oC
4	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	o = oQ
14	4	0	3	0	0	1	3	0	2	0	1	44	2	0	523	0	1	p = u
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	q = uC
2	0	0	1	0	0	0	0	0	0	0	0	4	2	0	33	0	0	r = uQ

b)
 === Matriz de Confusión ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	<-- clasificado
1466	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a = a
460	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = aC
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = aQ
735	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d = e
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	e = eC
66	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	f = eQ
322	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	g = ë
667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	h = ëC
141	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i = ëQ
629	0	0	62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	j = i
149	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k = iC
67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	l = iQ
225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	m = o
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n = oC
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	o = oQ
598	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p = u
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	q = uC
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	r = uQ

Fig.5.10. Estas matrices de confusión corresponden a la segunda serie de pruebas para la variante del distrito de Choapan. El inciso a pertenece al método MLP F1-5 y el inciso b al método MSV RBF F1-2.

En el Anexo L se encuentran las tablas completas que contienen los porcentajes de las instancias correctamente clasificadas según la base de datos, el método empleado para clasificarlas, el número de clases y atributos utilizados.

5.7. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).

A continuación se presentan los resultados obtenidos del método DTW que fue utilizado para determinar el símbolo fonético SAMPA de la vocal Ě y E que provienen de ambas base de datos.

Resultados del método DTW utilizando la herramienta Matlab

Los resultados obtenidos para la variante zapoteca de Choapan del estado Veracruz y la del distrito de Choapan indican que el símbolo SAMPA “8” es el que corresponde asignarle a la vocal Ě en cualquiera de sus tres realizaciones, como se observa en los resultados de las tablas 5.17 y 5.18 respectivamente. Además el símbolo se encuentra señalado mediante un círculo rojo dentro del trapecio de vocales IPA en la figura 5.11.

Tabla.5.17. Símbolos Fonéticos asignados a las vocales E y Ě en cualquiera de sus tres realizaciones para la variante de Choapan del estado de Veracruz.

Vocal	Símbolo		Palabra	Idioma	Género
	IPA	SAMPA			
Ě'	ə	8	buss ¹	Sueco	Femenino
Ě'Ě	ə	8	Buss	Sueco	Femenino
Ě	ə	8	Buss	Sueco	Femenino
E'	ɔ	O	ontem ²	Portugués	Femenino
E'E	ə	8	Buss	Sueco	Femenino
E	ə	8	Buss	Sueco	Femenino

Tabla.5.18. Símbolos Fonéticos asignados a las vocales E y Ě en cualquiera de sus tres realizaciones para la variante de Choapan del estado de Oaxaca.

Vocal	Símbolo		Palabra	Idioma	Genero
	IPA	SAMPA			
Ě'	ə	8	buss	Sueco	Femenino
Ě'Ě	ə	8	buss	Sueco	Femenino
Ě	ə	8	buss	Sueco	Femenino
E'	ə	8	buss	Sueco	Femenino
E'E	ə	8	buss	Sueco	Femenino
E	ə	8	buss	Sueco	Femenino

1.- La palabra sueca *buss* se traduce al español como: autobús.

2.- La palabra portuguesa *ontem* se traduce al español como: ayer.

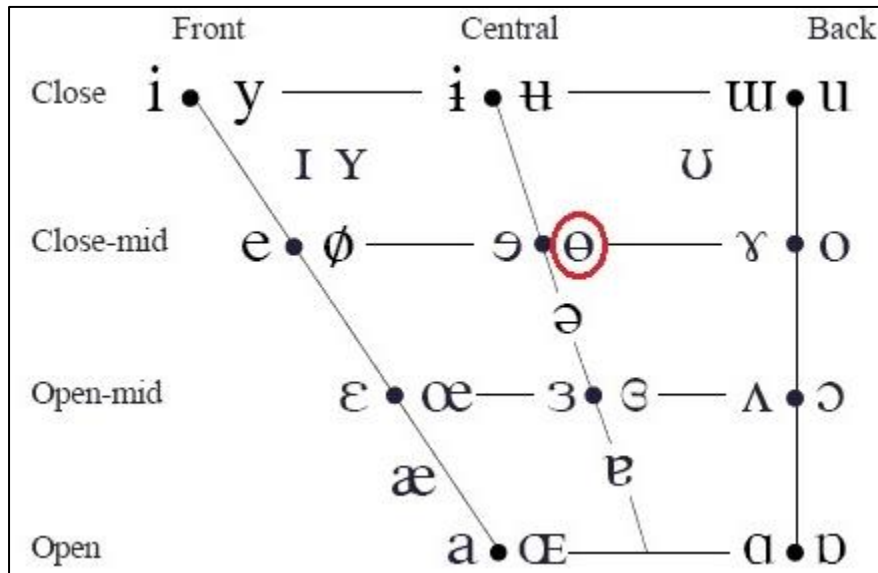


Fig.5.11. Trapecio de Vocales IPA resaltando el símbolo ə dentro de este.

Resultados del método DTW utilizando la herramienta Praat.

Los resultados obtenidos para la variante zapoteca de Choapan del estado Veracruz y la del distrito de Choapan indican que el símbolo SAMPA “E” es el que corresponde asignarle a la vocal Ę porque la mayoría de las realizaciones coinciden con este, como se observa en los resultados de las tablas 5.19 y 5.20 respectivamente. Además el símbolo se encuentra señalado mediante un círculo rojo dentro del trapecio de vocales IPA en la figura 5.12.

Tabla.5.19. Símbolos Fonéticos asignados a las vocales E y Ę en cualquiera de sus tres realizaciones para la variante de Choapan del estado de Veracruz.

Vocal	Símbolo		Palabra	Idioma	Genero
	IPA	SAMPA			
Ę'	ɛ	E	gesetz ³	Alemán	Femenino
Ę'Ę	ɛ	E	gesetz	Alemán	Femenino
Ę	ɛ	E	gesetz	Alemán	Femenino
E'	ɛ	E	gesetz	Alemán	Femenino
E'E	ɐ	6	besser ⁴	Alemán	Masculino
E	ɛ	E	gesetz	Alemán	Femenino

3.- La palabra alemana *gesetz* se traduce al español como: ley, código o sanción.

4.- La palabra alemana *besser* se traduce al español como: mejor, más o mejorado.

Tabla.5.20. Símbolos Fonéticos asignados a las vocales E y Ë en cualquiera de sus tres realizaciones para la variante de Choapan del estado de Oaxaca.

Vocal	Símbolo		Palabra	Idioma	Genero
	IPA	SAMPA			
Ë'	ə	@	ce ^s	Francés	Masculino
Ë'Ë	ɛ	E	gesetz	Alemán	Femenino
Ë	ɛ	E	gesetz	Alemán	Femenino
E'	ɛ	E	gesetz	Alemán	Femenino
E'E	ə	@	ce	Francés	Masculino
E	ɛ	E	gesetz	Alemán	Femenino

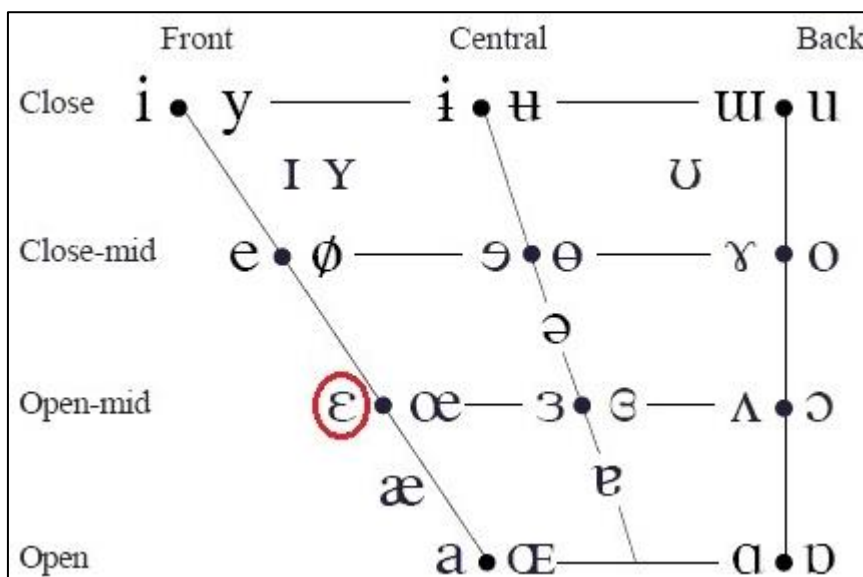


Fig.5.12. Trapecio de Vocales IPA resaltando el símbolo ε dentro de este.

5.- La palabra francesa *ce* se traduce al español como: este.

Los resultados anteriores se resumen en la tabla 5.21 donde se muestra según el software empleado (Matlab o Praat) el símbolo fonético obtenido para las vocales \ddot{E} y E en sus tres diferentes realizaciones, según la base de datos utilizada.

Tabla.5.21. Símbolos IPA y SAMPA asignados a las vocales \ddot{E} y E en sus tres realizaciones según los resultados anteriores.

Variante zapoteca del Distrito de Choapan.					Variante zapoteca de Choapan del estado de Veracruz				
Vocal	Resultado				Vocal	Resultado			
	Matlab		Praat			Matlab		Praat	
\ddot{E}	ə	ɤ	ɛ	E	\ddot{E}	ə	ɤ	ɛ	E
\ddot{E}'	ə	ɤ	ɛ	E	\ddot{E}'	ə	ɤ	ə	@
$\ddot{E}'\ddot{E}$	ə	ɤ	ɛ	E	$\ddot{E}'\ddot{E}$	ə	ɤ	ɛ	E
E	ə	ɤ	ɛ	E	E	ə	ɤ	ɛ	E
E'	ɔ	O	ɛ	E	E'	ə	ɤ	ɛ	E
E'E	ə	ɤ	ɐ	ɔ	E'E	ə	ɤ	ə	@

En el Anexo M se encuentra de forma desarrollada cada uno de los experimentos realizados en ambos programas de computadora.

6. Discusión de Resultados.

El estudio de las lenguas precolombinas presenta una serie de aspectos a tomar en cuenta a fin de desarrollar un análisis acústico adecuado y preciso. La lengua zapoteca se integra por 57 variantes donde el grado de inteligibilidad entre cada una de ellas representa un problema para la caracterización y modelado de los sonidos. Entre las variantes zapotecas se encuentra la de Choapan (zpc) que se habla principalmente en el Distrito de Choapan del estado de Oaxaca y también con una minoría de hablantes en el municipio de Playa Vicente, estado de Veracruz. Según la escala EGIDS señala que la variante zpc se encuentra en un nivel “8a”, es decir, en un estado moribundo de la lengua porque es utilizada en su mayoría por personas mayores y de la tercera edad mientras que los jóvenes la sustituyen por otras lenguas como el español.

El proyecto realizado se enfocó inicialmente en el análisis de la variante de Choapan del estado de Veracruz ya que se contaba con un hablante de esta región. Se tiene que resaltar que incluso el catálogo de lenguas propuesto por INALI desconoce a este lugar como una zona de habla zapoteca. Razón por la cual el interés para la realización del análisis se incrementa y a su vez también crea la necesidad de comparar los resultados de la región de Veracruz con los obtenidos para el distrito de Choapan de donde es originaria esta variante.

Al realizar el análisis de la lengua zapoteca se tiene que localizar la región de donde es el hablante, el tipo de variante a la que pertenece y el nivel en que se encuentra según la escala EGIDS. Cada uno de estos factores puede influir al conseguir tanto información ortográfica como gramatical además de afectar la construcción de la base de datos ya que entre más se encuentre en peligro de extinción la variante, se complica el hecho de conseguir individuos que cuenten con un alto grado de conocimientos y dominio sobre su lengua. Un ejemplo claro de esto se observó al analizar la variante de Choapan debido a que se encuentra en el estado moribundo de la lengua según la escala EGIDS, con lo que nos enfrentamos que al seleccionar un hablante joven su dominio sobre la lengua zapoteca se ha visto reducido por el uso del idioma español y cuando se desea comprobar la información obtenida se cuenta con número limitado de referencias bibliográficas.

La base de datos creada para este proyecto consiste de la grabación de un conjunto de 188 palabras que pertenecen a la Lista Swadesh, donde el hablante es una persona joven, originaria de la comunidad Arenal Santa Ana del municipio antes señalado del estado de Veracruz (en el proyecto se conoce como base de datos de la variante de Choapan del estado de Veracruz). La edad del hablante es un factor importante debido a que como señala la escala EGIDS, el hablante ha dejado de usar la lengua zapoteca por el español trayendo como consecuencia que no le resulte sencillo recordar la correcta pronunciación y ortografía de las palabras, por lo tanto se verificó esta última mediante la búsqueda de las palabras en diccionarios que pertenecen a la variante zpc o que tengan un alto grado de inteligibilidad como la variante Zoogocho (zpg).

Al realizar la búsqueda de las palabras nos encontramos con tres grados de dificultad, que influyen en cierta medida el análisis de la variante zpc y son:

- Dificultad nula. La palabra se encuentra en ambos diccionarios o principalmente en el que corresponde a la variante zpc.
- Dificultad intermedia. La palabra se encuentra igual en el diccionario de la variante zpg pero no así en el diccionario de la variante zpc, ya que ésta es total o parcialmente diferente a la propuesta por el hablante. La anterior discrepancia entre los diccionarios nos coloca un punto de alerta debido a que no nos podemos confiar en que las dos regiones antes mencionadas (la de Veracruz y la de Oaxaca) hablen la variante zpc, sino hay que tener presente que sólo existe un alto grado de inteligibilidad entre ellas.
- Dificultad alta. La palabra buscada no se encuentra o es totalmente diferente en ambos diccionarios. Esto nos lleva a tener que aplicar el método de reconocimiento a través de la similitud dada por DTW, para poder identificar algunos fonemas que participan dentro de las palabras.

Del conjunto total de palabras de la base de datos creada restan por identificar 42, lo que aunado al tamaño de la base de datos ocasiona que existan vocales que cuenten con una cantidad de ejemplos limitada, por lo que fue necesario utilizar el método DTW con la finalidad de encontrar el tipo de vocales que emplean las palabras desconocidas y así incrementar el número de ejemplos de éstas. Antes de utilizar el método DTW para encontrar la vocal desconocida, se realizó una prueba que permitió observar el comportamiento de este método sobre la base de datos con lo que encontramos que de un total de 50 instancias analizadas solo el 76% de ellas es reconocida de forma correcta.

La Lista Swadesh consiste de un conjunto de palabras que se consideran comunes en todas las lenguas y puede utilizarse como el corpus de la base de datos, debido a que al principio del proyecto no se tenían conocimientos relacionados a la lengua zapoteca. El beneficio de usar la lista se ve limitado por el hecho de que la base de datos creada tiene un desbalanceo fonético, es decir, para algunas vocales el número de ejemplos es alto mientras que otras tienen muy pocos o incluso nulos. El tamaño y el desbalanceo fonético de la base de datos creada provoca que se requiera una segunda base de datos con el fin de comprobar los resultados obtenidos en el análisis de vocales.

La segunda base de datos corresponde a la misma variante pero de la región de Oaxaca (en el proyecto se conoce como la base de datos del distrito de Choapan), se diferencia de la primera en que es de mayor tamaño beneficiando el número de ejemplos para cada vocal lo que contrarresta el desbalanceo fonético que ésta posee. Comparando el número de vocales en ambas bases de datos se tiene que las dos coinciden en que las vocales con mayor cantidad de ejemplos son “A, E, I” en su realización sencilla. En lo que respecta a las vocales con el menor número de sonidos ambas bases difieren: para la base de datos de la variante de Choapan del estado de Veracruz las vocales son: O’ (O cortada) y U’U (U quebrada), mientras que para la base de datos del distrito de Choapan son: O’O (O quebrada) y U’ (U cortada).

Para realizar el análisis de las bases de datos fue necesario extraer cada uno de los fonemas realizando una segmentación automática mediante el software HTK, el cual entrega los límites de los fonemas para posteriormente ser extraídos los audios mediante un script en el programa Praat. La exactitud con que determinaron los límites en ambas bases de datos fue relativamente buena para tratarse de una lengua que no es común de segmentarse dentro de HTK, en el caso de la base de datos de la variante de Choapan del estado de Veracruz al ser de un tamaño reducido pudo verificarse manualmente cada uno de los límites de los fonemas los cuales no necesitaron ser modificados. Por otro lado, la base de datos del distrito de Choapan debido a que contiene 8034 fonemas de vocales resulta complicado revisar cada uno de los límites, así que la estrategia para medir la calidad de éstos fue a través de la duración de las vocales. Proponiendo que valores por debajo de 60ms no resultan propios de una vocal y pueden considerarse como errores en el proceso de segmentación, de tal forma que solo el 72.07% de los fonemas fueron segmentados correctamente ya que cumplen con la condición de duración.

Una factor que pudo ser determinante en el resultado de la segmentación automática es la velocidad con que se grabaron las palabras en la base de datos; la que corresponde a la variante de Choapan del estado de Veracruz sus palabras se pronunciaron con aproximadamente la misma velocidad, además antes y después de cada palabra se realiza una pausa. La base de datos de la variante del distrito de Choapan almacena oraciones que contienen palabras pronunciadas a diferentes velocidades y utiliza pausas desiguales entre las palabras, es decir, varía el tamaño de la separación entre cada una de ellas.

Se extrajeron las características acústicas o espectrales de cada uno de los sonidos de las 18 vocales que se encuentran en ambas bases de datos con el propósito de diferenciar entre los tipos de vocales (A, E, È, I, O, U) o el total de ellas. La primera serie de análisis consistió en graficar las formantes F1 contra F2 (F1 vs. F2) con lo que se observaron claramente los seis grupos que corresponden a los tipos de vocal, incluso se apreció la similitud entre las vocales E y È debido a la cercanía entre los dos conjuntos de fonemas. Además se buscó identificar mediante la gráfica F1 contra F2 las dieciocho vocales obteniendo así un resultado no satisfactorio debido a que no se forma un agrupamiento claro para alguna de ellas, incluso si se hubiera tomado de forma separada cada tipo de vocal con sus tres realizaciones. También se constató que la cantidad de datos no influye en los resultados de las gráficas.

Con el fin de encontrar una diferencia que permitiera identificar entre las realizaciones de las vocales, se realizó una gráfica de caja de las siguientes características: intensidad, duración y frecuencia fundamental. Mediante estas gráficas se observó que para ambas bases de datos el traslape de cada una de las cajas no ofreció la suficiente variación para identificar alguna de las dieciocho vocales. También al igual que la gráfica F1 vs. F2 el tamaño de la base de datos no influyó en los resultados.

En el proyecto las formantes fueron las mejores características al momento de realizar la identificación de las vocales, como pudo constatarse tanto en la representación gráfica de las mismas o mediante el método Perceptrones Multicapa. Éste obtuvo un porcentaje alto de instancias que se clasificaron correctamente y el cual es aceptable si se compara con resultados obtenidos en trabajos previos como en la

referencia [22], donde se realiza el reconocimiento de vocales sobre una lengua tonal utilizando MLP obteniendo resultados que varían desde 90.93% (aplicado al conjunto de datos de entrenamiento) hasta un 67.73 % (aplicado al conjunto de datos de prueba).

Para ambas bases de datos los porcentaje de la cantidad de sonidos de vocales correctamente clasificados fueron más altos cuando se utilizaron los 6 tipos de vocales como clases y disminuyeron significativamente cuando se reconocieron 18 clases, probablemente ocurrió porque la configuración del tracto vocal para cada una de las realizaciones es aproximadamente la misma y por lo tanto los valores de las formantes utilizadas para identificar las vocales se asemejan. Por otro lado las Maquinas de Soporte Vectorial utilizando un kernel RBF obtuvieron el mayor porcentaje de instancias incorrectamente clasificadas sin importar la cantidad de clases utilizadas y si observamos la matriz de confusión nos percatamos que este método agrupa a la mayoría de las instancias analizadas en la clase de la vocal A, posiblemente por ser la que tiene la mayor cantidad de datos.

Para determinar el símbolo fonético SAMPA de la vocal \ddot{E} se buscó la semejanza acústica entre ésta en cualquiera de las tres realizaciones contra ejemplos de sonidos de vocales que se encuentran dentro del trapecio IPA (no olvidando la correspondencia con el alfabeto SAMPA), que fueron extraídos de cualquier lengua del mundo que tenga una transcripción fonética y que utilice alguno de estos símbolos. Las comparaciones entre los sonidos se realizaron con el método DTW mediante la utilización de diferente software, obteniendo los siguientes resultados:

- Matlab. El símbolo fonético IPA asignado a esta vocal es “ø” (equivale al símbolo “8” en SAMPA) que se obtuvo de la segmentación del sonido en la palabra sueca “buss” (autobús).
- Praat. El símbolo fonético IPA asignado a la vocal es “ɛ” (equivale al símbolo “E” en SAMPA) segmentado de la palabra alemana “gesetz” (ley), aunque para la realización cortada de la variante del distrito de Choapan el símbolo fonético IPA asignado fue “ə” o “@” en SAMPA que fue extraído de la palabra francesa “ce” (este).

El método DTW fue utilizado en dos partes dentro de la metodología: para “determinar fonemas” y para “determinar el símbolo fonético SAMPA de la vocal \ddot{E} ”. La primera utiliza los audios de prueba y prototipo de una misma lengua mientras en la segunda los audios de prueba corresponden a la lengua zapoteca en tanto que los prototipos corresponden a diferentes idiomas. De acuerdo a lo anterior DTW no encuentra diferencias significativas entre la vocal \ddot{E} y E, ya que al comparar los símbolos asignados a estas vocales en la mayoría de los casos son muy similares, independientemente del método utilizado. En lo que respecta a las realizaciones sucede algo similar a lo anterior para la mayoría de los casos, con excepción del análisis en Praat de la vocal E’E (E quebrada) que asignó un símbolo diferente al obtenido en las otras realizaciones, aunque consultando las tablas de resultados para ambas bases de datos en el Anexo M, se puede observar que el segundo símbolo que más se asemeja coincide con el asignado a las otras realizaciones para la base de datos de mayor tamaño, mientras que la de menor tamaño cuenta con un número de ejemplos muy reducido para

esta vocal lo cual puede influir en el resultado. Además si se observan dos símbolos resultantes para la vocal E'E en el trapecio para ambas bases de datos se nota que son muy semejantes y sólo cambia el tamaño de la abertura de la boca.

En el proyecto las formantes fueron las mejores características al momento de realizar la identificación de las vocales, como pudo constatarse tanto en la representación gráfica de las mismas o mediante el método Perceptrones multicapa, el cual obtuvo un porcentaje alto de instancias que se clasificaron correctamente. La metodología utilizada en el análisis de las vocales fue construyéndose a lo largo de la realización del proyecto, es decir, cada uno de los pasos que se implementaron surgió con base en las necesidades que se requerían. Por ejemplo, cuando se buscaba obtener el conjunto de fonemas para las vocales de la base de datos se utilizó la "segmentación automática" o al presentarse problemas para los sonidos de algunas palabras se realizó la "determinación de fonemas".

7. Conclusiones y Trabajo Futuro.

Se realizó un análisis de las señales de voz de una variante en peligro de extinción del zapoteca. Dada esta situación el análisis se desarrolló en dos bases de datos que fueron obtenidas a través de la grabación de un hablante nativo y de la obtención de señales de voz previamente grabadas de otra región que comparte la misma variante.

El análisis se enfocó en las vocales y sus variantes por lo que se pudo obtener ventajas con el uso de las características acústicas de las señales, en particular de las formantes. Las realizaciones de las 6 vocales se pueden observar con facilidad a través de la aplicación de las formantes, pero el efecto no se presenta al tratar de identificar las realizaciones normal, cortada y quebrada de las mismas. La diferenciación entre las vocales E y Ę se realizó a partir de DTW, el cual permitió comparar el sonido en cuestión con sonidos caracterizados en otros idiomas. En este sentido, se nota una semejanza con sonidos del sueco, que es una lengua indoeuropea con características tonales.

Al clasificar las vocales del zapoteco se encontró que el porcentaje de clasificación correcta disminuye cuando se analizan las 18 realizaciones con respecto a las 6 vocales en general. Aún así, el menor valor está por arriba del 50% cuando se utilizan las formantes como características y una red neuronal como sistema de clasificación.

Este proyecto significó un avance dentro del análisis de la variante de Choapan de la lengua zapoteca, éste se realizó en un tipo específico de sonidos (fonemas de las vocales). Falta desarrollar el análisis de las consonantes con el fin de encontrar las características que permitan identificarlas. También se tiene que mejorar la segmentación de estos sonidos ya sea utilizando el HTK, mediante la modificación de los parámetros de configuración a los valores que mejor se adapten a las velocidades que utilizan los hablantes al pronunciar oraciones o palabras, o haciendo uso de algún otro método o sistema que garantice una segmentación precisa. Con respecto a los clasificadores se puede mejorar el porcentaje de instancias correctamente identificadas mediante el incremento de atributos (características extraídas) en el análisis; el uso de características como los coeficientes cepstrales puede ser de utilidad para estos fines siempre que se tenga un número importante de ejemplos por sonido.

Por el momento la base de datos sólo pertenece a la variante zapoteca de Choapan así que como trabajo posterior se buscarán hablantes de otras variantes de la lengua con el fin de analizar comparativamente el efecto de las realizaciones de los sonidos así como el efecto de la tonalidad en el zapoteca. Asimismo se espera aumentar el tamaño de la base de datos de la variante de Choapan y procurar que ésta sea fonéticamente balanceada para tener suficiente información que de lugar a un análisis más robusto de esta variante coadyuvando a su conservación.

Anexos.

A. Código ISO639.3 y la escala EGIDS de las variantes zapotecas.

El estándar ISO639.3 consiste de un código de tres letras que se le asigna a todas las lenguas del mundo, de tal forma que la lengua zapoteca se identifica con las letras “zap”. Cada una de las 57 variantes que la integran también tiene su propio código ISO, el cual se encuentra en la siguiente tabla; se incluye el estado y nivel según la escala EGIDS que tiene cada variante.

Tabla.A.1. Código ISO639.3 y Escala EGIDS de las variantes zapotecas.

No.	Variante	Código ISO639.3	Escala EGIDS	
			Nivel	Estado
1	Aloápam	zaq	6a	Vigoroso
2	Amatlán	zpo	5	Desarrollo
3	Asunción Mixtepec	zpo	8b	Cerca de la Extinción
4	Ayoquesco	zaf	6b	Amenazado
5	Cajonos	zad	5	Desarrollo
6	Chichicapan	zpv	7	Cambiando
7	Choapan	zpc	8a	Moribundo
8	Coatecas Altas	zca	6a	Vigoroso
9	Coatlán	zps	7	Cambiando
10	El Alto	zpp	6b	Amenazado
11	Elotepec	zte	7	Cambiando
12	Guevea de Humboldt	zpg	5	Desarrollo
13	Güilá	ztu	6a	Vigoroso
14	Istmo	zai	5	Desarrollo
15	Lachiguiri	zpa	6b	Amenazado
16	Lachixío	zpl	5	Desarrollo
17	Lapaguía - Guivini	ztl	6a	Vigoroso
18	Loxicha	ztp	6a	Vigoroso
19	Mazaltepec	zpy	6a	Vigoroso
20	Miahuatlán	zam	6b	Amenazado
21	Mitla	zaw	5	Desarrollo

22	Mixtepec	zpm	5	Desarrollo
23	Ocotlán	zac	6a	Vigoroso
24	Ozolotepec	zao	5	Desarrollo
25	Petapa	zpe	6b	Amenazado
26	Quiavicuzas	zpj	6a	Vigoroso
27	Quioquitani - Quierí	ztq	5	Desarrollo
28	Rincón	zar	5	Desarrollo
29	San Agustín Mixtepec	zpx	8b	Cerca de Extinción
30	San Baltazar Loxicha	zpx	6b	Amenazado
31	San Juan Guelavía	zab	5	Desarrollo
32	San Pedro Quiatoni	zpf	5	Desarrollo
33	San Vicente Coatlán	zpt	6a	Vigoroso
34	Santa Catarina Albarradas	ztn	6b	Amenazado
35	Santa Inés Yatzechi	zpn	6a	Vigoroso
36	Santa María Quiegolani	zpi	5	Desarrollo
37	Santiago Xanica	zpr	6b	Amenazado
38	Santo Domingo Albarradas	zas	5	Desarrollo
39	Sierra de Juárez	zaa	6a	Vigoroso
40	Rincón del Sur	zsr	5	Desarrollo
41	Ixtlán del Sudeste	zpd	6b	Amenazado
42	Tabaa	zat	6a	Vigoroso
43	Tejalapan	ztt	8a	Moribundo
44	Texmelucan	zpz	5	Desarrollo
45	Tilquiapan	zts	6a	Vigoroso
46	Tlacolulita	zpk	8a	Moribundo
47	Totomachapan	zph	7	Cambiando
48	Xadani	zax	7	Cambiando
49	Xanaguía	ztg	6a	Vigoroso
50	Yalálag	zpu	5	Desarrollo
51	Yareni	zae	6a	Vigoroso
52	Yatee	zty	5	Desarrollo

53	Yatzachi	zav	5	Desarrollo
54	Yautepec	zpb	7	Cambiando
55	Zaachila	ztx	7	Cambiando
56	Zaniza	zpw	6b	Amenazado
57	Zoogocho	zpq	5	Desarrollo

B. Zapoteco de Veracruz.

La lengua zapoteca se habla principalmente en el estado de Oaxaca aunque actualmente hay registros de su existencia en otros estados de la República Mexicana como México y Veracruz, este último se integra por 212 municipios [58] de los cuales sólo en Minatitlán, Coatzacoalcos, Santiago Sochiapan y Playa Vicente se habla alguna variante de la lengua zapoteca. En específico para estos dos últimos municipios se hablan las variantes de Villa Alta y Choapan [59], en la figura B.1 se encuentra el mapa de la región geográfica donde se hablan estas variantes, y en la tabla B.1 muestra las comunidades de los municipios de Veracruz donde se utilizan éstas.

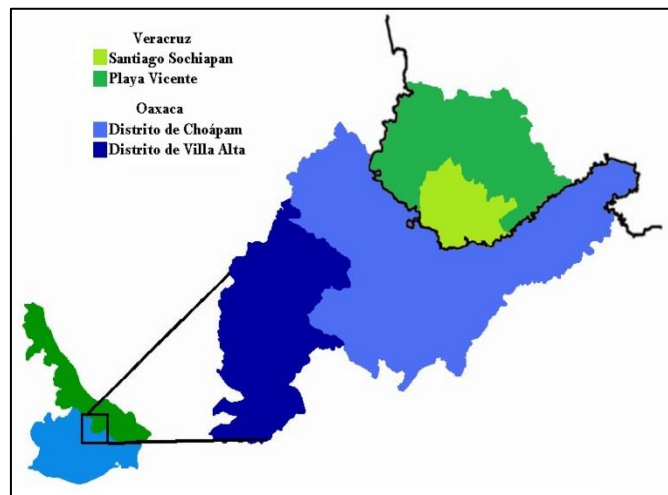


Fig.B.1. Mapa que indica la zona geográfica donde se hablan las variantes de Choapan y Villa Alta.

Tabla. B.1. Comunidades de los municipios de Playa Vicente y Santiago Sochiapan del estado de Veracruz donde se hablan las variantes de Choapan y Villa Alta.

Variante	Municipio	Comunidades
Choapan	Playa Vicente	El Arenal Santa Ana, Chilapa del Carmen, Ampliación Chilapa, Barrio Las Cruces, El Nigromante, El Zapotal, Arroyo Zacate, Arroyo Dehesa, El Tomatillo, Zanja de Caña, El Mirador, Arroyo Guadalupe La Palma, Arroyo Colorado
	Santiago Sochiapan	Tatahuicapa, Niños Héroe, Boca del Monte, Santa Teresa, Xochiapan, Nuevo Boca
Villa Alta	Playa Vicente	El Nigromante, Carmen Arroyo, El Gavilán
	Santiago Sochiapan	Ejido Revolución, Lázaro Cárdenas

C. Playa Vicente y la Comunidad Arenal Santa Ana.

Playa Vicente. [60]

Geografía.

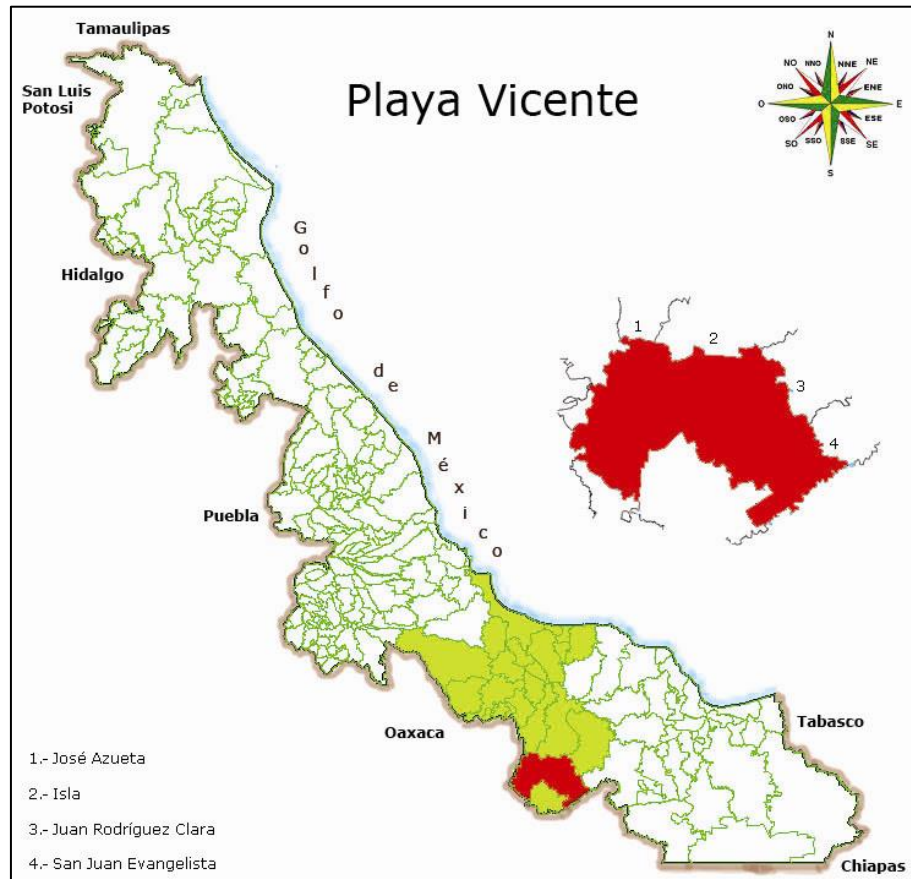


Fig.C.1. Mapa del estado de Veracruz, donde se indica el lugar donde se encuentra el municipio de Playa Vicente y sus colindancias con otros municipios.

Ubicación.

Entre los paralelos 17° 34' y 17° 57' de latitud norte; los meridianos 95° 21' y 95° 54' de longitud oeste; altitud entre 10 y 200 m.

Límites.

Colinda al norte con el estado de Oaxaca y los municipios de José Azueta, Isla y Juan Rodríguez Clara; al este con los municipios de Juan Rodríguez Clara, San Juan Evangelista y el estado de Oaxaca; al sur con el estado de Oaxaca y el municipio de Santiago Sochiapan; al oeste con el municipio de Santiago Sochiapan y el estado de Oaxaca.

Clima.

- Cálido subhúmedo con lluvias en verano (74%) y cálido húmedo con abundantes lluvias en verano (26%).

- Rango de temperatura 24 – 28°C
- Rango de precipitación 1 400 – 2 600 mm

DATOS GEOGRÁFICOS	
Indicador	Valor
Cabecera municipal	Playa Vicente
Localidades en 2010	4
Urbanas	252
Rurales	248
Superficie	1,174.2 km ²
Porcentaje del territorio estatal	1.6%
Densidad poblacional en 2010	34.9 hab/km ²

Fuente: SEFIPLAN con datos de INEGI, Censo de Población y Vivienda 2010 y Marco Geoestadístico Municipal 2010; y Gobierno del Estado de Veracruz, Plan Veracruzano de Desarrollo 2011/2016.

Demografía.

EVOLUCIÓN DE LA POBLACIÓN				
Año	Total	Hombres	Mujeres	Proporción estatal (%)
2010	40,984	19,660	21,324	0.54
2005	38,125	18,062	20,063	0.54
2000	49,388	23,938	25,450	0.71
1995	52,754	26,468	26,286	0.78
1990	49,579	24,990	24,589	0.80
1980	51,231	26,247	24,984	0.95

Fuente: INEGI. Censos y Conteos de Población y Vivienda, 1980 a 2010.

HABITANTES EN PRINCIPALES LOCALIDADES, 2010	
Localidad	Habitantes
Playa Vicente	9,083
Nuevo Ixcatlán	3,770
Abasolo del Valle	3,691
El Nigromante	2,594
Arenal Santa Ana	1,551
Resto de localidades	20,295

Fuente: INEGI. Censo de Población y Vivienda 2010.

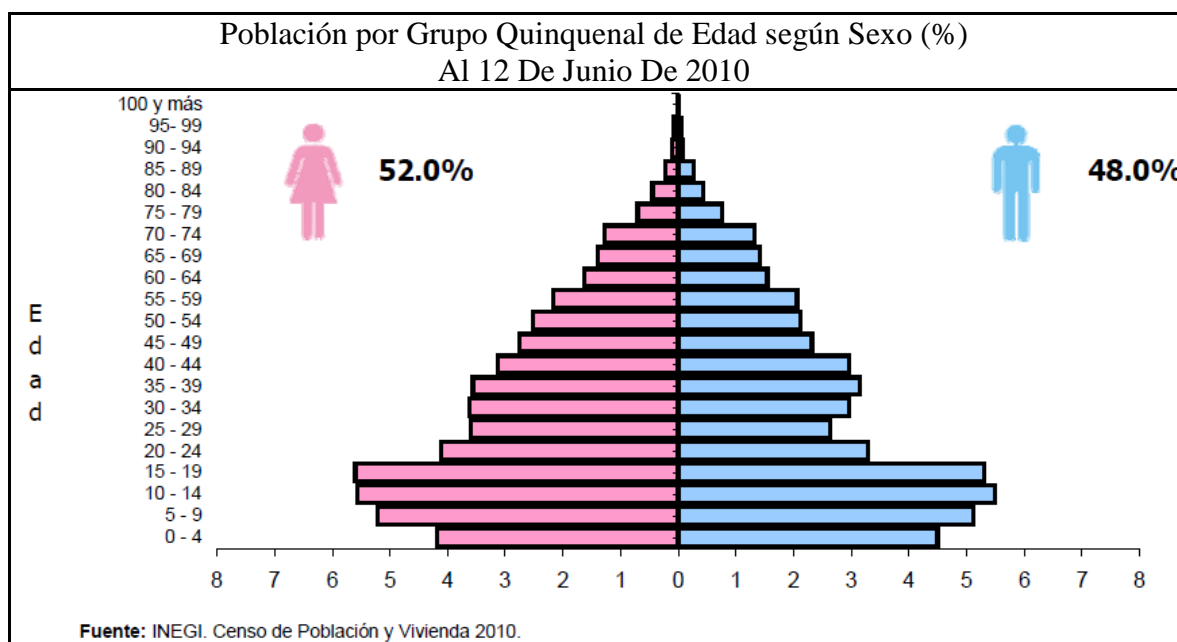
POBLACIÓN POR TAMAÑO DE LOCALIDAD, 2010	
Ámbito	Habitantes
Tamaño	
Rural	21,846

Menos de 500 habitantes	14,140
500 a 2,499 habitantes	7,706
Urbano	19,138
2,500 a 14,999 habitantes	19,138
15,000 y más habitantes	0

Fuente: INEGI. Censo de Población y Vivienda 2010.

EDAD MEDIANA, 2010	
Indicador	Valor
Edad mediana (años)	25
Hombres	25
Mujeres	26

Fuente: INEGI. Censo de Población y Vivienda 2010.



Población por Grupos de Edad, 2010	
Grupo de edad	Habitantes
Infantil (0-14 años)	12,327
Joven y adulta (15-64 años)	25,011
Tercera edad (65 años y más)	3,525

NOTA: Excluye a la población que no especificó su edad, por lo que la suma puede no coincidir con el total de población expresado en el cuadro de la evolución de la población.

Fuente: INEGI. Censo de Población y Vivienda 2010.

Población Indígena, 2010	
Indicador	Valor

Población en hogares indígenas a/	17,652
Población de 3 años o más hablante de lengua indígena	
Total	10,113
Hombres	4,706
Mujeres	5,407
Población de 3 años y más que habla lengua indígena	26.10%
Hablantes de lengua indígena que no hablan español	3.66%
Lengua principal	Mazateco

a/ Se refiere a la población en hogares donde el jefe (a) o su cónyuge habla alguna lengua indígena.

Fuente: INEGI. Censo de Población y Vivienda 2010.

Desarrollo Social.

Analfabetismo, 2010	
Indicador	Valor
Población de 6 a 14 años que sabe leer y escribir	83.5%
Población de 15 años y más	28,536
Población de 15 años y más analfabeta	5,285
Tasa de analfabetismo	18.7%

Fuente: INEGI. Censo de Población y Vivienda 2010.

Comunidad Arenal Santa Ana.

Información de localidad. [8]

Datos actuales	
Nombre de la Entidad	Veracruz de Ignacio de la Llave
Nombre del Municipio	Playa Vicente
Grado de marginación municipal 2010	Alto
Nombre de la localidad	Arenal Santa Ana

Datos demográficos						
Año	2005			2010		
	Hombres	Mujeres	Total	Hombres	Mujeres	Total
Total de población en la localidad	699	829	1,528	720	831	1,551
Viviendas particulares habitadas	348			373		
Grado de marginación de la	Alto			Alto		

localidad		
Grado de rezago social localidad	Bajo	Bajo

Ubicación.

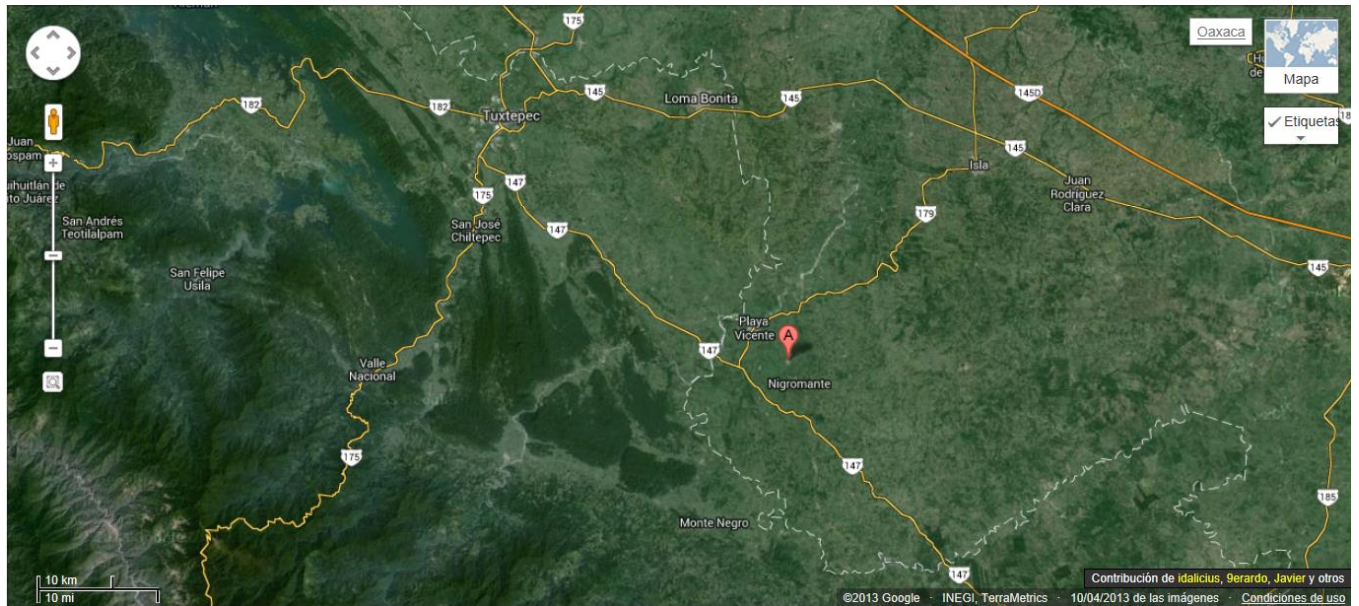


Fig.C.2. Mapa de la ubicación de la Comunidad Arenal Santa Ana municipio de Playa Vicente, Veracruz. Se encuentra indicada en el punto A del mapa. [61]

D. Lista Swadesh.

La Lista Swadesh es una herramienta lingüística que consiste de un conjunto de palabras que se consideran comunes en todas las lenguas, inicialmente estaba integrada por 100 palabras pero en la actualidad la lista tiene un tamaño más grande, para el caso del proyecto se utiliza una lista de 188 palabras [62] la cual se muestra en las siguientes tablas. Estas se componen por un par de columnas; la primera indica la palabra en español y la segunda proporciona su traducción al zapoteco de Choapan.

La primera parte de la Lista Swadesh se encuentra en la tabla D.1 y está compuesta de aquellas palabras en zapoteco donde la ortografía proporcionada por el hablante de la variante de Choapan del estado de Veracruz coincide con la que se encuentra en los diccionarios de las variantes de Choapan o Zoogocho que se hablan en el estado de Oaxaca [10, 40, 41]. La segunda parte de la lista Swadesh se encuentra en la tabla D.2 y está compuesta por las palabras en zapoteco cuya ortografía es proporcionada por el hablante solamente debido a que ésta no pudo corroborarse en ningún diccionario.

Tabla.D.1. Parte de la Lista Swadesh donde la ortografía de las palabras en zapoteco fue corroborada en los diccionarios.

No.	Español	Zapoteco
1	afilado	tuzi'
2	agua	nisa
3	ahí	lena'
4	ala	xila
5	amarillo	gachi
6	ancho	laga'
7	animal	bia
8	año	guiza
9	aquí, acá	cani'la
10	árbol	yaga yula
11	barriga	le'e
12	beber	ra'a
13	boca	ra'
14	bosque	gui'shidau
15	bueno	dyi'a
16	cabello	guitza'
17	cabeza	guichoro
18	caminar	sa'
19	camino	neza

No.	Español	Zapoteco
74	levantarse	chasa
75	lleno	nža
76	lluvia	nisayo
77	luna	biu'
78	madre	xna'
79	malo	ben xhni'a
80	mano	na'a
81	mar	nis'adau
82	matar	gutia
83	mojado	bisa
84	morir	gatia
85	muchos	bacaside'
86	mujer	gula
87	nariz	xquirú
88	negro	gaso
89	niño	xcuidi'
90	no	o'o
91	noche	bagula
92	nombre	lauru

20	cantar	gula
21	carne	bě'ëla
22	cenizas	de
23	cerca	ga'la'
24	cielo	guiebaa
25	cinco	ga'yu'n
26	cola	xpam
27	comer	gahua
28	como	nacan
29	contar	laban
30	corazón	lužtau'
31	correcto	sia
32	cortar	choga
33	corto	dadau'zi
34	cuando	bataga
35	cuatro	tapan
36	cuerda	du
37	decir	nian
38	delgado	lacu
39	día	teza
40	diente	leya'
41	donde	ga'n
42	dormir	rasia
43	dos	chupan
44	él	lëubi'
45	ellos	leuyacabi
46	empujar	chži'gan
47	escupir	re'en
48	ese, aquel	bina'
49	espalda	excu'zu'
50	esposa	gula quia
51	esposo	bě'm byu quia
52	este	daní
53	estrecho	gui
54	estrella	belo
55	flor	yo'o
56	frío	zaga

93	nuevo	cubi
94	ojo	lahua'
95	oreja	naga'
96	otro	itun
97	padre	xuza
98	pájaro	guinni
99	palo	yaga
100	partir	la'an
101	pecho	chu
102	pelear	dila
103	pensar	pensare
104	perro	bécu'
105	pesado	zi'i
106	pez, pescado	bela
107	pie	ni'a
108	piedra	yo
109	piel	guidi
110	piojo	bechi'
111	pluma	dú'bi'
112	poco	lë'tin
113	podrido	guzu
114	porque	bin
115	que	bida'n
116	quemar	rei
117	quien	nu'
118	raíz	lun
119	recto	li'i
120	río	yau
121	rojo	žna
122	saber	neziu
123	sal	zede'
124	sangre	ren
125	seco	bidyi
126	semilla	le'xi
127	sentarse	rue
128	serpiente	bela
129	sol	ubiza

57	fuego	gui'bela
58	girar	chechoga
59	grande	že
60	grasa	za
61	gusano	bela
62	hierba	guixi'
63	hoja	la'ga'
64	hombre	bë'm byu
65	hueso	chita
66	huevo	chíta
67	humano	benë'
68	humo	dzen
69	izquierda	guegala
70	jugar	guito
71	largo	tuna
72	lejos	tu'na
73	lengua	luzi

130	temer	chžeba
131	tierra	yu
132	tirar	chu'nán
133	todo	yogun
134	tres	tzonan
135	tripas	xlo
136	tu	lë'
137	uno	tun
138	uña	xucu'
139	ustedes	lëe'
140	venir	guelá
141	ver	le'eda'
142	verde	ya'a
143	viejo	dagula
144	viento	be
145	vivir	zacaza
146	yo	në'di'

Tabla.D.1. Parte de la Lista Swadesh donde la ortografía de las palabras en zapoteco pertenece a la proporcionada por el hablante.

No.	Español	Zapoteco
1	agacharse	cuegue
2	algunos	nulan na'
3	apretar	quichan
4	apuñalar	quiba
5	arañar	iniza
6	arena	roo
7	atar	schogan
8	blanco	chigichi
9	caer	gezo
10	cálido	szla
11	cazar	raguaba
12	chupar	zupia
13	coser	guein
14	cuello	gueru
15	cuerno	luzubaá
16	dar	huia

No.	Español	Zapoteco
22	lago	elgun
23	lavar	giva
24	limpiar	stia
25	liso	zili
26	montaña	gizidau
27	morder	riña'
28	niebla	lanza
29	nosotros	neturu
30	nube	chulá
31	oír	rea'
32	oler	nupian
33	pierna	zun
34	polvo	baestei
35	reír	riya'
36	respirar	reziabe
37	rodilla	rarriba

17	golpear	izana
18	gordo	mallí
19	hígado	guirri
20	hincharse	gi
21	jalar	zan

38	soplar	luba
39	sucio	guchi
40	tener	quia
41	volar	saa
42	vomitarse	gueba

E. Transcripción fonética de la base de datos de la variante de Choapan del estado de Veracruz.

La transcripción fonética para la Segmentación con HTK consistió en buscar un símbolo fonético que representase a cada uno de los sonidos de la lengua zapoteca en su variante de Choapan. Las reglas utilizadas para hacer la transcripción se encuentran dentro del paso dos de la metodología, aplicándolas se obtuvo la transcripción de la tabla E.1 para las palabras de la base de datos de Choapan del estado de Veracruz.

Tabla.E.1. Transcripción Fonética de las palabras de la base de datos de la variante de Choapan del estado de Veracruz.

No.	Palabra	Transcripción
1	BACASIDE'	b a k a s i D e C
2	BAGULA	b a G u l a
3	BATAGA	b a t a G a
4	BE	b e
5	BECHI'	b e t S i C
6	BECU'	b e k u C
7	BELA	b e l a
8	BELO	b e l o
9	BEN	b e n
10	BENĒ'	b e n N C
11	BESTE	b e K t e
12	BIA	b j a
13	BIDA'N	b i D a C n
14	BIDYI	b i d y i
15	BIN	b i n
16	BINA'	b i n a C
17	BISA	b i s a
18	BIU'	b j u C
19	BYU	b j j u
20	BĒ'M	b N C m
21	BĒ'ĒLA	b N Q l a
22	CAN'LA	k a n i C l a
23	CHASA	t S a s a
24	CHECHOGA	t S e t S o G a
25	CHIGUICHI	t S i G i t S i
26	CHITA	t S i t a

No.	Palabra	Transcripción
93	LENA'	l e n a C
94	LEUYACAB I	l e w j j a k a B i
95	LEYA'	l e j j a C
96	LII'	l i Q
97	LUBA	l u B a
98	LUN	l u n
99	LUZI	l u T i
100	LUZTAU'	l u Z t a u C
101	LUZUBAA	l u T u B a a
102	LĒ'	l N C
103	LĒ'TIN	l N C t i n
104	LĒE'	l N E
105	LĒUBI'	l N w B I
106	MALLI	m a L i
107	NA'	n a C
108	NA'A	n a Q
109	NACAN	n a k a n
110	NAGA'	n a G a C
111	NETURU	n e t u r u
112	NEZA	n e T a
113	NEZIU	n e T j u
114	NIA'	n i C a
115	NIAN	n j a n
116	NISA	n i s a
117	NISA'DAU	n i s a C D a w
118	NISAYO	n i s a j j o

27	CHOGA	tS o G a
28	CHS'IA	tS K iCa
29	CHU	tS u
30	CHU'NAN	tS uC n a n
31	CHULA	tS u l a
32	CHUPAN	tS u p a n
33	CHZEBA	tS Z e B a
34	CHZI'GAN	tS Z iC G a n
35	CUBI	k u B i
36	CUEGUE	k w e G e
37	DADAU'ZI	d a D a uC T i
38	DAGULA	d a G u l a
39	DANI	d a n i
40	DE	d e
41	DILA	d i l a
42	DU	d u
43	DU'BI'	d uC B iC
44	DY'A	d y iCa
45	DZEN	d z e n
46	ELGUN	e l g u n
47	EXCU'ZU'	e G s k uC T uC
48	GA'LA'	g aC l aC
49	GA'N	g aC n
50	GA'YU'N	g aC j j uC n
51	GACHI	g a tS i
52	GAHUA	g a w a
53	GASO	g a s o
54	GATIA	g a t j a
55	GUCHI	g u tS i
56	GUEBA	g e B a
57	GUEGALA	g e G a l a
58	GUEIN	g e j n
59	GUELA	g e l a
60	GUERU	g e r u
61	GUI	g i
62	GUI'BELA	g iC B e l a
63	GUI'SHIDAU	g iC s i D a w
64	GUIBA	g i B a
65	GUICHORO	g i tS o r o
66	GUIDI	g i D i

119	NU'	n uC
120	NULAN	n u l a n
121	NUPIAN	n u p j a n
122	NZA	n Z a
123	NĒ'DI'	n NC d iC
124	O'O	oQ
125	PENSARE	p e n s a r e
126	QUIA	k j a
127	QUIBA	k i B a
128	QUICHAN	k i tS a n
129	RA'	r aC
130	RA'A	r aQ
131	RAGUABA	r a G w a B a
132	RARRIBA	r a r r i B a
133	RASIA	r a s j a
134	RE'EN	r eQ n
135	REA'	r eA
136	REI	r r e j
137	REN	r r e n
138	REZIABE	r r e T j a B e
139	RIYA'	r r i j j aC
140	RIÑA'	r r i J aC
141	ROO	r r o o
142	RUE	r r w e
143	SA'	s aC
144	SAA	s a a
145	SCHOGAN	s tS o G a n
146	SIA	s j a
147	SZLA	s K l a
148	TAPAN	t a p a n
149	TEZA	t e T a
150	TU'NA	t uC n a
151	TUN	t u n
152	TUNA	t u n a
153	TUZI'	t u T iC
154	TZONAN	t T o n a n
155	UBIZA	u B i T a
156	XCUIDI'	K s k w j D iC
157	XHNI'A	G s n iCa
158	XILA	K s i l a

67	GUIEBAA	g je B aa
68	GUINNI	g i n n i
69	GUIRRI	g i r r i
70	GUIITO	g i t o
71	GUITZA'	g i DT aC
72	GUIXI'	g i Gs iC
73	GUIZA	g i T a
74	GUIZIDAU	g i T i D aw
75	GULA	g u l a
76	GUTIA	g u t j a
77	GUZU	g u T u
78	HUIA	w j a
79	INIZA	i n i T a
80	ITUN	i t u n
81	IZANA	i T a n a
82	LA'AN	l a Q n
83	LA'GA'	l aC G aC
84	LABAN	l a B a n
85	LACU	l a k u
86	LAGA'	l a G aC
87	LAHUA'	l a w aC
88	LANZA	l a n T a
89	LAURU	l a w r u
90	LE'E	l e Q
91	LE'EDA'	l e Q D aC
92	LE'XI	l eC Gs i

159	XLO	Ks l o
160	XNA'	Gs n aC
161	XPAM	Ks p a m
162	XQUIRU	Ks k i r u
163	XUCU'	Ks u k uC
164	XUZA	Ks u T a
165	YA'A	jj aQ
166	YAGA	jj a G a
167	YAU	jj aw
168	YO	jj o
169	YO'O	jj oQ
170	YOGUN	jj o G u n
171	YU	jj u
172	YULA	jj u l a
173	ZA	T a
174	ZACAZA	T a k a T a
175	ZAGA	T a G a
176	ZAN	T a n
177	ZE	Z e
178	ZEDE'	T e D eC
179	ZI'I	T i Q
180	ZILI	T i l i
181	ZNA	Z n a
182	ZUN	T u n
183	ZUPIA	T u p j a

F. Pasos de la segmentación HTK.

La segmentación automática se realiza a través del software HTK (Hidden Markov Model Toolkit); éste consiste de un conjunto de herramientas que permiten construir y manipular los modelos ocultos de Markov. El proceso de segmentación se divide en dos etapas: entrenamiento y segmentación. La primera consiste en entrenar los modelos para cada uno de los fonemas que participan dentro de los audios y la segunda etapa es la encargada de obtener los límites de los fonemas. Pero antes de iniciar el proceso de segmentación es necesario preparar los audios, carpetas y archivos que se utilizan.

Los pasos y scripts ocupados durante el proceso de segmentación están basados en los que se encuentran en la referencia [39] haciendo pequeñas modificaciones de acuerdo a las necesidades del proyecto.

Preparación de los datos.

Audios.

HTK manipula varios formatos de audio entre los que se encuentran el tipo wav, que es el utilizado en este proyecto. Para mayor información sobre las características que deben de tener los audios se puede consultar la referencia [37].

Organización.

El organizar los archivos a través de carpetas es un punto esencial de la segmentación HTK porque facilita su manejo y comprensión. Las carpetas utilizadas se muestran en la tabla F.1.

Tabla.F.1. Carpetas cómo se organiza la información para el manejo de HTK.

Carpeta	Contenido
Data	Contiene un archivo (extensión txt) que almacena la transcripción de los audios además se incluyen los archivos de transcripción uno por cada audio (extensión lab).
Dbase	La base de datos de los audios a segmentar
Scr	Incluye los scripts y archivos complementarios que se crean antes o durante el proceso de segmentación, por lo regular tienen extensiones como: mlf, list, scp, cfg, list, dict y dicc.
Models	Integra las subcarpetas con los modelos hmm, una por cada reestimación realizada

Transcripción de los audios.

Para cada archivo de audio que se encuentra en la base de datos a segmentar se realiza una transcripción la cual se almacena con el mismo nombre que el archivo de audio, pero con la extensión “lab” como se muestra en los ejemplos de la tabla F.2.

Tabla.F.2. Ejemplos del contenido de los archivos de transcripción.

Archivo (.wav)	Transcripción	
	Archivo (.lab)	Contenido
Ala1.wav	Ala1.lab	XILA
capIIIparrflaa.wav	capIIIparrflaa.lab	UZU TU BEN' PARTIDO FARISEO LAONĚ' NICODEMO

También se puede usar un archivo de texto denominado transcripciones.txt donde su formato consiste del nombre del audio seguido de la transcripción como se muestra en las siguientes líneas:

```
Ala1 XILA
capIIIparrflaa UZU TU BEN' PARTIDO FARISEO LAONĚ'
NICODEMO
```

Diccionario principal.

El diccionario principal es un conjunto de palabras de la lengua con su respectiva transcripción fonética, éste archivo generalmente se nombra como “dicc.dict” y se localiza en la carpeta scr. El diccionario debe contener como mínimo todas las palabras que aparecen en los audios, ordenadas alfabéticamente y sin repeticiones con el fin de evitar posteriores errores en el proceso de segmentación. En la figura F.1 se muestra el formato que debe tener un diccionario donde por default se agregan: los dos renglones del principio y el último.

!ENTER	[] sil
!EXIT	[] sil
A	a
ABDYIN	a B dy i n
ABEDILANDO'	a B e D i l a n D oC
ABELOLE	a B e l o l e
ABEYALO	a B e j j a l o
ABRAHAM	a B r a a m
.	
.	
ZĚ	T N
ZĚ'Ě	T NQ
ZĚDI	T N D i
ZĚNĚ'	T N n NC
ZĚRA	T N r a
sil	[] sil

Fig.F.1. Fragmento del diccionario principal (dicc..dict”).

Etapa de Entrenamiento.

1.- Formando la lista de palabras.

Se construye la lista de las palabras que se encuentren dentro de los audios, esto se puede realizar de dos formas: mediante la instrucción *cut* en donde las palabras están ordenadas alfabéticamente o por medio de la instrucción *cat* que obtiene la lista pero las palabras se encuentran desordenadas. La forma en cómo se realiza esto se encuentra en las tablas F.3 y F.4 respectivamente para cada instrucción, además en la figura F.2 se muestra una parte de cada una de estas listas.

Tabla.F.3. Forma de obtener la lista de palabras con la instrucción *cut*.

Instrucción	<code>cut -d ' ' -f2- ../data/transcripciones.txt tr '\040' '\012' sort uniq > word.list</code>
Herramienta HTK	
Script utilizado	
Archivo utilizado	transcripciones.txt
Resultado	Se muestra en la figura F.2 inciso “a”
Dirección destino	zapsegm/scr/word.list

Tabla.F.4. Forma de obtener la lista de palabras con la instrucción *cat*.

Instrucción	<code>cat ../data/*.lab ./wordList > word.list</code>
Herramienta HTK	
Script utilizado	wordList
Archivo utilizado	Todos los archivos con extensión “lab”
Resultado	Se muestra en la figura F.2 inciso “b”
Dirección destino	zapsegm/scr/word.list

a)	b)
BÃ< 'Ã< LA	CUEGUE
BÃ< 'Ã< LA	CUEGUE
BÃ< 'MYU	TUZI '
BÃ< 'MYU	TUZI '
BÃ< 'NE '	NISA
BÃ< 'NE '	NISA
BACACIDE	LENA '
BACACIDE	LENA '
BATAGA	XILA
BATAGA	XILA
BECHII '	NULAN
BECHI '	NA '
BECU '	NULAN
BE:CU '	NA '
BELA	

Fig.F.2. Listas de palabras que aparecen en las transcripciones. El inciso a) se obtiene con la instrucción cut y el inciso b) con la instrucción cat.

2.- Creando el diccionario secundario.

El diccionario secundario proporciona las palabras que se encuentran en los audios junto con la transcripción fonética de las mismas agregando al final de cada renglón el símbolo de pausa corta (sp), además las palabras están ordenadas alfabéticamente y sin repeticiones. El diccionario puede realizarse manualmente aunque la manera más fácil de obtenerlo es a través de la herramienta HDMan que proporciona (ver tabla F5):

- El diccionario secundario con el nombre “dict.list” (un fragmento del diccionario se muestra en la figura F.3).
- El archivo phone1.list que indica el conjunto de fonemas utilizados en las transcripciones
- El archivo “dlog” que proporciona un resumen; el número de veces que se repite cada fonema en el documento, la cantidad de palabras analizadas, las palabras que se encuentran repetidas o las que no fueron encontradas dentro del diccionario principal (palabras desconocidas).

Tabla.F.5. Creación del diccionario secundario con la instrucción HDMan.

Instrucción	HDMan -m -w word.list -n phone1.list -l dlog dict.list dicc.dict
Herramienta HTK	HDMan
Script utilizado	global.ded
Archivo utilizado	word.list, dicc.dict
Resultado	phone1.list, dict.list, dlog
Dirección destino	zapsegm/scr/phone1.list zapsegm/scr/dict.list zapsegm/scr/dlog.list

```

BACACIDE      b a k a T i D e sp
BATAGA        b a t a G a sp
BECHI '       b e t S i sp
BECU '        b e k u sp
BELA          b e l a sp
BIA           b j a sp
B\303\213 'MYU b N m j j u sp
:
:
TUNA          t u n a sp
TZONAN        t T o n a n sp
VIDA 'N       b i D a n sp
XCUIDI '      G s k w i D i sp
XNA '         G s n a sp

```

Fig.F.3. Fragmento del diccionario secundario dict.list.

Para mayor información sobre el comando HDMan se puede consultar el manual HTKBook [37].

3.- Creamos dos listas de fonemas utilizados.

En el proceso de segmentación se necesitan dos listas de fonemas que se crean a partir de la lista “phone1.list” (obtenida en el paso anterior). La primera incluirá el símbolo de silencio (sil) en tanto que la segunda contendrá el símbolo de silencio (sil) y el de pausas cortas (sp). Los pasos para obtenerlas son:

3.A. Se crea una lista que no contiene el símbolo sp con la instrucción grev como se muestra en la tabla F.6.

Tabla.F.6. Forma de crear la lista sin el símbolo de pausas cortas (sp).

Instrucción	grev -v sp phone1.list > phone0.list
Herramienta HTK	
Script utilizado	
Archivo utilizado	phone1.list
Resultado	phone0.list
Dirección destino	zapsegm/scr/phone0.list

3.B. Agregar el símbolo de silencio (sil) a los archivos phone0.list y phone1.list (ver la tabla F.7).

Tabla.F.7. Forma de agregar el símbolo de silencio (sil).

Instrucción	echo sil>> phone0.list echo sil>> phone1.list
Herramienta HTK	
Script utilizado	
Archivo utilizado	phone0.list, phone1.list
Resultado	phone1.list, phone0.list
Dirección destino	zapsegm/scr/phone0.list zapsegm/scr/phone1.list

4.- *Transcripciones de los archivos de audio (archivos mlf).*

Creamos dos tipos de archivos mlf (Master Label File) que contienen la transcripciones a nivel palabra o fonema de cada uno de los audios.

4.A. Transcripción a nivel palabra. Se crea un archivo mlf que contendrá las palabras que posee cada archivo “lab”. Esto se realiza de forma automática mediante el script: “pts2mlf.pl”, la forma en como se utiliza se encuentra en la tabla F.8. Un fragmento del archivo resultante se muestra en la figura F.4.

Tabla.F.8. Aplicación del script para obtener la transcripción a nivel palabra.

Instrucción	pts2mlf.pl word0.mlf transcripciones.txt
Herramienta HTK	
Script utilizado	pts2mlf.pl
Archivo utilizado	transcripciones.txt
Resultado	word0.mlf
Dirección destino	zapsegm/scr/word0.mlf

```

#!MLF!#
"*/capIparrf1a.lab"
CAORA
BINECUE'
LAO
YEDYI
LAYU
BAZUCAZI
TU
BEN'
NACA
XTI'IDZA'
DIUZI
LĚBEN'
NA'
ZULĚNĚ'
DIUZI
TUZ?E
LĚSCAN'
NAQUĚ'
DIUZI
.
"*/capIparrf1b.lab"
DESDE
AODIXU
DIUZI
YEDYI
LAYU
BAZULANĚ'

```

Fig.F.4. Transcripción ortográfica a nivel palabra.

4.B. Transcripción a nivel fonema. Para realizar la transcripción de los archivos “lab” fonema por fonema se utiliza la herramienta HLEd como se encuentra en la tabla F.9 y una parte del contenido del archivo se muestra en la figura F.5.

Tabla.F.9. Aplicación del script para obtener la transcripción a nivel palabra.

Instrucción	HLEd -l '*' -d dict.list -i phone0.mlf mkPhones0.led word0.mlf
Herramienta HTK	HLEd
Script utilizado	mkPhones0.led
Archivo utilizado	dict.list, word0.mlf
Resultado	phone0.mlf
Dirección destino	zapsegm/scr/phone0.mlf

```

#!MLF!#
"*/capIparrfla.lab"
sil
k
ao
r
a
b
i
n
k
weC
l
ao
jj
e
dy
.
.
i
sil
.
"*/capIparrflb.lab"
sil
d

```

Fig.F.5. Fragmento del archivo mlf creado para la transcripción a nivel fonema.

5.- Se obtiene una lista de los archivos de audio a segmentar.

Para obtener la lista de los audios que se van a segmentar se utiliza el script copyList como se encuentra en la tabla F.10.

Tabla.F.10. Implementación del script copyList.

Instrucción	./copyList > train.scp
Herramienta HTK	
Script utilizado	copyList
Archivo utilizado	
Resultado	train.scp
Dirección destino	zapsegm/scr/train.scp

6.- Parametrización de las señales de audio.

Se parametriza las señales de audio en secuencias de características de acuerdo a la configuración proporcionada en el archivo “comp.cfg” que contiene: el tipo de MFCC a utilizar, el formato de audio, el tipo de ventana, el tamaño de la misma, el número de coeficientes, etc. Los valores iniciales de la media y la varianza se proporcionan mediante el archivo “proto” el cual es reevaluado y almacenado en la carpeta hmm0. Todos estos archivos se le proporcionan a la herramienta HCompV como se muestra en la tabla F.11.

Tabla.F.11. Parametrización de las señales.

Instrucción	HCompV -C comp.cfg -f 0.01 -m -S train.scp -M ../models/hmm0 proto
Herramienta HTK	HCompV
Script utilizado	proto, comp.cfg
Archivo utilizado	train.scp
Resultado	proto, vFloors, hmmdefs
Dirección destino	zapsegm/models/hmm0/vFloors zapsegm/models/hmm0/proto zapsegm/models/hmm0/hmmdefs

7.- Creación del archivo “macros”.

Se obtiene mediante la concatenación del archivo “vFloors” y el templete “mactmp”, éste último define: las opciones globales, el tipo de parámetro HMM y el tamaño del vector. La creación del archivo macros se realiza mediante la instrucción cat como en la tabla F.12.

Tabla.F.12. Creando archivo macros.

Instrucción	cat mactmp.txt ../models/hmm0/vFloors > ../models/hmm0/macros
Herramienta HTK	
Script utilizado	
Archivo utilizado	mactmp.txt, vFloors
Resultado	macros
Dirección destino	zapsegm/models/hmm0/macros

8.-Se reproduce el modelo HMM inicial para cada uno de los fonemas.

Realizamos una copia de “proto” para cada uno de los fonemas que participan en las transcripciones (como se encuentra en la tabla F.13), incluyendo el símbolo del silencio (sil). Se utiliza la lista de fonemas que se encuentran en el archivo “phone0.list” (no contiene el símbolo sp).

Tabla.F.13. Define los valores para el modelo HMM de cada fonema.

Instrucción	<code>./init phone0.list > ../models/hmm0/hmmdefs</code>
Herramienta HTK	
Script utilizado	Init
Archivo utilizado	phone0.list
Resultado	Hmmdefs
Dirección destino	zapsegm/models/hmm0/hmmdefs

9.- Primeras reestimaciones.

Se realizan las primeras tres reestimaciones de los modelos HMM utilizando la herramienta HERest, la forma en como se introduce la instrucción en HTK se encuentra en la figura F.14. El resultado de una reestimación se almacena en una nueva carpeta hmm.

Tabla.F.14. Primeras reestimaciones de los modelos HMM.

Instrucción	<code>HERest -C comp.cfg -I phone0.mlf -t 250.0 150.0 1000.0 -S train.scp -T 1 -H ../models/hmm0/macros -H ../models/hmm0/hmmdefs -M ../models/hmm1 phone0.list</code>
Herramienta HTK	HERest
Script utilizado	
Archivo utilizado	comp.cfg, phone0.mlf, train.scp, phone0.list macros y hmmdefs (de la carpeta hmm0)
Resultado	hmmdefs y macros
Dirección destino	zapsegm/models/hmm1/macros zapsegm/models/hmm1/hmmdefs

10.- Integrando el modelo HMM inicial para las pausas cortas (sp).

A partir de ahora se inicia el análisis con pausas cortas modificando los archivos “hmmdefs y macros”, los cuales se almacenan en una nueva carpeta (hmm4). La forma de obtener éstos es:

10.A. El archivo “hmmdefs” se crea mediante el script “sil2sp” que realiza una copia de los modelos HMM de cada uno de los fonemas, donde se incluye el modelo HMM inicial para sp y en la tabla F.15 se muestra como se realiza esto.

Tabla.F.15. Creando el modelo HMM para la pausa corta (sp).

Instrucción	<code>./sil2sp ../models/hmm3/hmmdefs > ../models/hmm4/hmmdefs</code>
Herramienta HTK	
Script utilizado	sil2sp
Archivo utilizado	hmmdefs (de la carpeta hmm3)
Resultado	Hmmdefs
Dirección destino	zapsegm/models/hmm4/hmmdefs

10.B. El archivo macros se copia de la carpeta hmm anterior, la instrucción de como se realiza esto se encuentra en la tabla F.16.

Tabla.F.16. Copiando el archivo macros.

Instrucción	cp ../models/hmm3/macros ../models/hmm4/macros
Herramienta HTK	
Script utilizado	
Archivo utilizado	macros (de la carpeta hmm3)
Resultado	Macros
Dirección destino	zapsegm/models/hmm4/macros

11. Enlazando el modelo HMM de la pausa corta.

Enlazamos los modelos HMM para “sp” y “sil” obteniendo un diagrama de estados como el que se muestra en la página 34 del HTKBook [37]. La herramienta utilizada para hacer este enlace es HHed, la forma en como se utiliza se encuentra en la tabla F.17.

Tabla.F.17. Enlazando los modelos HMM para sp y sil.

Instrucción	HHed -H ../models/hmm4/macros -H ../models/hmm4/hmmdefs -M ../models/hmm5 sil.hed phone1.list
Herramienta HTK	HHed
Script utilizado	sil.hed
Archivo utilizado	macros y hmmdefs (de la carpeta hmm4) phone1.list
Resultado	hmmdefs, macros
Dirección destino	zapsegm/models/hmm5/hmmdefs zapsegm/models/hmm5/macros

12.- Transcripción mlf a nivel fonema con sp.

Se obtiene una transcripción mlf a nivel fonema utilizando las pausas cortas entre cada una de las palabras que contiene el audio, esto se realiza empleando la herramienta HLEd. En la tabla F.18 se encuentra la forma en como se utiliza esta herramienta y un fragmento del archivo obtenido se muestra en la figura F.6.

Tabla.F.18. Creando el archivo mlf utilizando pausa corta (sp).

Instrucción	HLEd -l ‘*’ -d dict.list -i phone1.mlf mkPhones1.led word0.mlf
Herramienta HTK	HLEd
Script utilizado	mkPhones1.led
Archivo utilizado	dict.list, word0.mlf
Resultado	phone1.mlf
Dirección destino	zapsegm/scr/phone1.mlf

```

#!MLF!#
"/capIparrfla.lab"
0 600000 sil sil
600000 1200000 k CAORA
1200000 2800000 ao
2800000 3200000 r
3200000 5000000 a
5000000 5000000 sp
5000000 6100000 b BINECUE'
6100000 6400000 i
6400000 7700000 n
7700000 8900000 k
8900000 10100000 weC
10100000 10100000 sp
10100000 10800000 l LAO
10800000 11600000 ao
11600000 11600000 sp
.
.
.
86400000 87500000 d DIUZI
87500000 89200000 jw
89200000 90000000 T
90000000 91400000 i
91400000 91900000 sp
91900000 92300000 sil sil

```

Fig.F.6. Archivo mlf de las transcripciones de los audios que incluye las pausas cortas (sp).

13.- Reestimaciones con pausas cortas

Se termina la etapa de entrenamiento realizando dos reestimaciones, utilizando la herramienta HERest como se encuentra en la tabla F.19.

Tabla.F.19. Restimaciones utilizando las pausas cortas (sp).

Instrucción	HERest -C comp.cfg -I phone1.mlf -t 250.0 150.0 1000.0 -S train.scp -T 1 -H ../models/hmm5/macros -H ../models/hmm5/hmmdefs -M ../models/hmm6 phone1.list
Herramienta HTK	HERest
Script utilizado	
Archivo utilizado	comp.cfg, phone1.mlf, train.scp, phone1.list macros y hmmdefs (de la carpeta hmm5)
Resultado	hmmdefs y macros
Dirección destino	zapsegm1/models/hmm6/macros zapsegm1/models/hmm6/hmmdefs

Etapa de segmentación.

14.- Agregando el símbolo de silencio (sil) al diccionario secundario (dict.list).

En la tabla F.20 se encuentra la instrucción para agregar el símbolo sil al diccionario dict.list.

Tabla.F.20. Agregando el símbolo sil al diccionario dict.list.

Instrucción	echo "sil sil" >> dict.list
Herramienta HTK	
Script utilizado	
Archivo utilizado	dict.list
Resultado	dict.list con la línea agregada sil sil
Dirección destino	

15.- Primera realineación de los datos.

La realineación de los datos se obtiene mediante la herramienta HVite donde los datos almacenados en la carpeta hmm7 pasan de una transcripción de palabras a fonemas, considerando todas las pronunciaciones de cada palabra y las salidas que mejor se parezcan a las señales de voz. La forma en como se utiliza esta herramienta se ilustra en la tabla F.21.

Tabla.F.21. Realineación de los datos utilizando la herramienta HVite.

Instrucción	HVite -l '*' -o SWT -b sil -C comp.cfg -a -H ../models/hmm7/macros -H ../models/hmm7/hmmdefs -i phone2.mlf -m -t 250.0 -y lab -I word0.mlf -S train.scp dict.list phone1.list
Herramienta HTK	HVite
Script utilizado	
Archivo utilizado	hmmdefs y macros (de la carpeta 7) dict.list, word0.mlf, comp.cfg, train.scp, phone1.list
Resultado	phone2.mlf
Dirección destino	zapsegm/scr/phone2.mlf

16.- Ultimas reestimaciones.

Realizamos otras dos reestimaciones para establecer nuevamente los parámetros hmm, la carpeta final será hmm9. Para realizar las últimas reestimaciones se utiliza la herramienta HERest como se encuentra en la tabla F.22.

Tabla.F.22. Ultimas reestimaciones.

Instrucción	HERest -C comp.cfg -I phone2.mlf -t 250.0 150.0 1000.0 -S train.scp -T 1 -H ../models/hmm7/macros -H ../models/hmm7/hmmdefs -M ../models/hmm8 phone1.list
Herramienta HTK	HERest
Script utilizado	
Archivo utilizado	comp.cfg, phone2.mlf, train.scp, phone1.list macros y hmmdefs (de la carpeta hmm7)
Resultado	hmmdefs y macros
Dirección destino	zapsegm1/models/hmm8/macros zapsegm1/models/hmm8/hmmdefs

17.- Segunda realineación de los datos.

La segunda realineación de los datos es el último paso de la etapa de segmentación y se realiza con el objetivo de obtener las pronunciaciones que más se parezcan a las señales de audio. Los resultados de los límites de los fonemas y palabras segmentados se encuentran en el archivo "word1.mlf". Esta realineación se realiza por medio de la herramienta HVite como se muestra en la tabla F.23.

Tabla.F.23. Realineación de los datos para obtener la segmentación.

Instrucción	HVite -l '*' -o SM -b sil -C comp.cfg -a -H ../models/hmm9/macros -H ../models/hmm9/hmmdefs -i word1.mlf -m -t 250.0 -y lab -I word0.mlf -S train.scp dict.list phone1.list
Herramienta HTK	HVite
Script utilizado	
Archivo utilizado	hmmdefs y macros de la carpeta 9 dict.list, word0.mlf, comp.cfg, train.scp, phone1.list
Resultado	word1.mlf
Dirección destino	zapsegm/scr/word0.mlf

18.- Creación de los archivos TextGrid.

Se crea un archivo TextGrid por cada uno de los audios de la base de datos segmentada. Los archivos TextGrid pueden ser de dos tipos: palabras o fonemas.

- TextGrid de palabras. Indica los límites de cada palabra que contiene el audio, se obtienen por medio del script wordLine como se muestra en la tabla F.24. En la figura F.7 se muestra el formato de dos de estos archivos.
- TextGrid de fonemas. Indica en un solo archivo los límites tanto de palabras como fonemas que contiene el audio, se obtiene por medio del script textGrid como se muestra en la tabla F.25. En la figura F.8 se muestra el formato de dos de estos archivos.

Tabla.F.24. Se crea de forma automática la carpeta wordgrids que contendrá la segmentación por palabra de cada archivo de audio.

Instrucción	./wordLine word1.mlf
Herramienta HTK	
Script utilizado	wordLine
Archivo utilizado	word1.mlf
Resultado	Se muestra un ejemplo de estos archivos en la Figura E.10
Dirección destino	zapsegm/scr/wordgrids/...

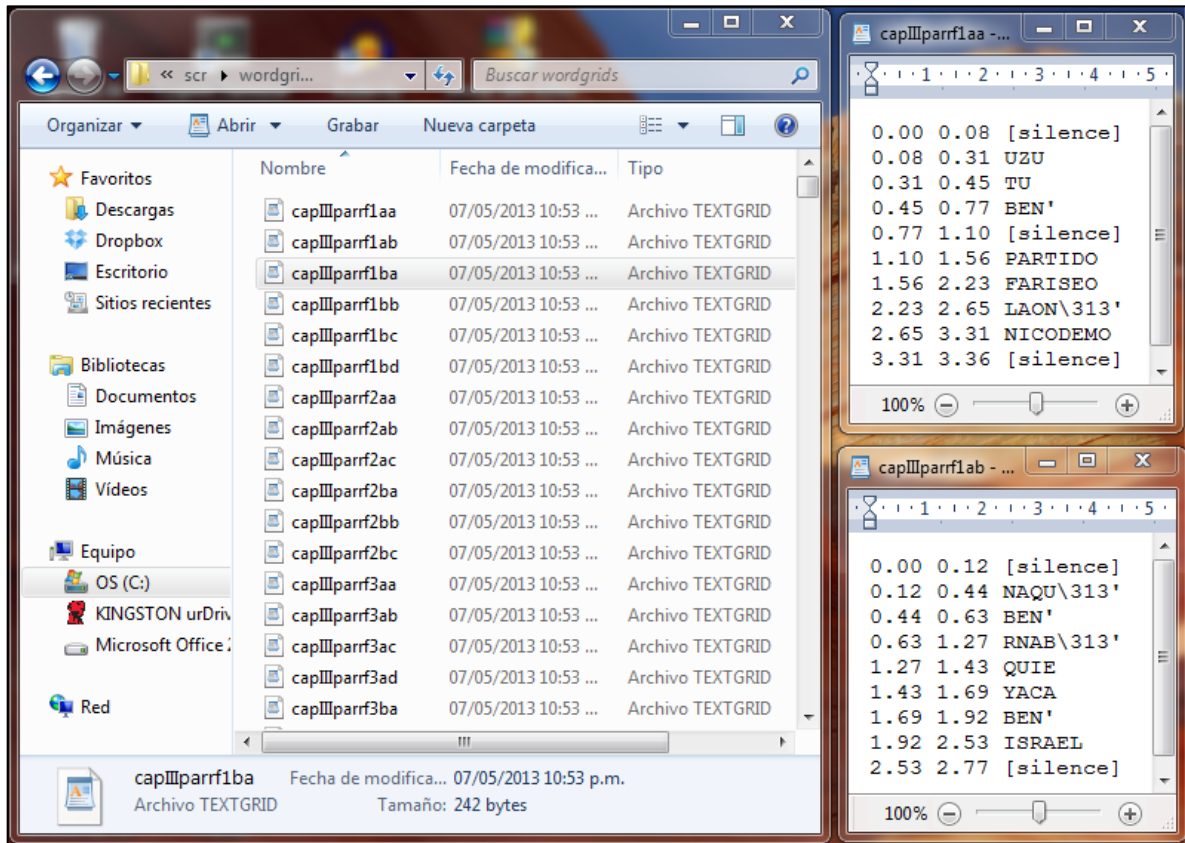


Fig.F.7. Ejemplos del contenido de los archivos TextGrid para la segmentación de las palabras.

Tabla.F.25. Se crea de forma automática la carpeta Fonemas que contendrá la segmentación de cada archivo de audio por fonema que lo integra.

Instrucción	./textGrid word1.mlf
Herramienta HTK	
Script utilizado	textGrid
Archivo utilizado	word1.mlf
Resultado	Se muestra un ejemplo de estos archivos en la Figura E.11
Dirección destino	zapsegm/scr/Fonemas/...

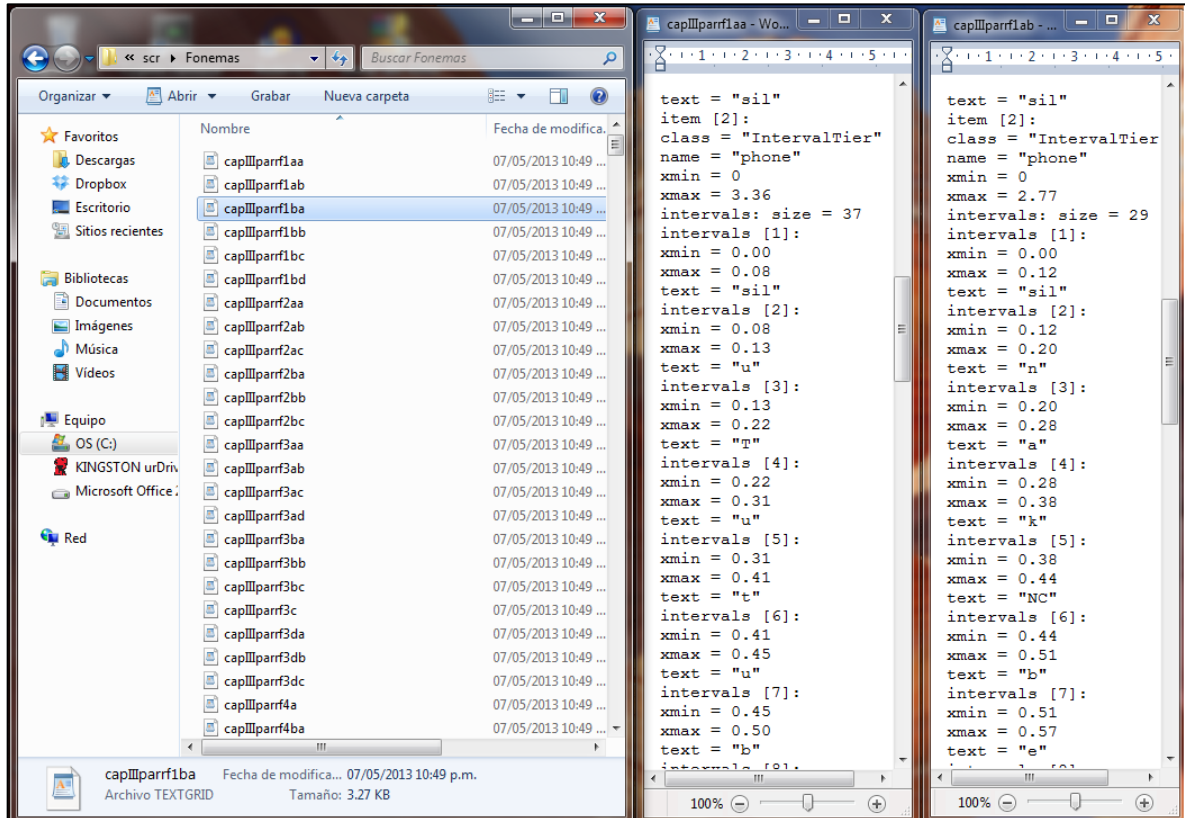


Fig.F.8. Ejemplos del contenido de los archivos TextGrid para la segmentación de los fonemas.

G. Alineamiento Temporal Dinámico (Dynamic Time Warping, DTW)

La técnica DTW consiste en encontrar la alineación óptima entre dos secuencias (que pueden ser dependientes del tiempo) considerando ciertas restricciones, donde la alineación busca que éstas se igualen. El campo de aplicación de esta técnica va desde la minería de datos hasta el reconocimiento de voz, en este último las señales de voz se convierten en señales discretas para ser comparadas.

DTW realiza la comparación de las características de cada una de las secuencias mediante la utilización de una medida de costos, que permite construir una “matriz de costos”, la cual es empleada por DTW para calcular una trayectoria dentro de ésta cuyo costo sea mínimo y que por lo tanto indique la semejanza entre ambas señales. Esta técnica puede realizarse en dos etapas: la construcción de la matriz de costos y el cálculo de la trayectoria óptima.

Construyendo la matriz de costos.

Supongamos que tenemos dos secuencias de características: $X = (x_1, x_2, \dots, x_N)$; $N \in \mathbb{N}$ y $Y = (y_1, y_2, \dots, y_M)$; $M \in \mathbb{N}$, donde el espacio de características estaría dado por F tomando en cuenta que $(x_n, y_m) \in F$ para $n \in [1: N]$ y $m \in [1: M]$. Para comparar ambas secuencias de características X y Y se necesita una medida de costo local, la cual debe cumplir con la siguiente condición: $c: F \times F \rightarrow \mathbb{R}_{\geq 0}$

El costo local se calcula por medio de distancias como pueden ser: Euclidiana, Manhattan, Mahalanobis, Cosenos, etc. De esta forma los costos indican la semejanza entre características, es decir, para valores pequeños la similitud entre las ellas es mayor e inversamente, entre mayor sea el costo más diferentes son. Se evalúa cada par de características de las secuencias X y Y obteniendo así la matriz de costos $c \in \mathbb{R}^{N \times M}$ definida por $c(n, m) = c(x_n, y_m)$ para que posteriormente DTW encuentre una alineación dentro de ésta que tenga el mínimo costo total.

Calculo de la trayectoria óptima.

La trayectoria encontrada dentro de la matriz que tenga el menor costo total se puede considerar como una trayectoria Warping si cumple con tres condiciones: límite, que sea monótona y el tamaño del paso. A continuación se muestra en qué consisten las condiciones.

Suponiendo a la trayectoria Warping como una secuencia de la forma $P = (p_1, \dots, p_L)$ con $p_l = (n_l, m_l) \in [1: N] \times [1: M]$ para $l \in [1: L]$, (N y M son los límites de la matriz de costos) se satisface las siguientes condiciones:

- Límite: $p_1 = (1, 1)$ y $p_L = (N, M)$
- Monótona: $n_1 < n_2 < \dots < n_L$ y $m_1 < m_2 < \dots < m_L$
- Tamaño del paso: $p_{l+1} - p_l \in \{(1, 0), (0, 1), (1, 1)\}$ para $l \in [1: L - 1]$

El costo total $C_P(X, Y)$ de una trayectoria Warping P entre las secuencias comparadas X y Y se puede expresar mediante la medida de costo local c de la siguiente forma: $C_P = \sum_{l=1}^L c(x_{n_l}, y_{m_l})$

Para una matriz de costos pueden existir varias trayectorias Warping pero no necesariamente todas tienen el mismo costo total, así que la técnica DTW debe detectar cuál de ellas tiene el mínimo valor. A este tipo de trayectoria se denomina como la trayectoria óptima Warping (P^*) y para determinarla se debe probar todas las posibles trayectorias Warping entre X y Y .

La trayectoria óptima Warping (P^*) se puede definir como la distancia $DTW(X,Y)$ entre X y Y donde el costo total P^* es:

$$DTW(X,Y) = C_P(X,Y) \\ = \min\{C_P(X,P) | P \text{ es una } (N,M) \text{ trayectoria Warping}\}$$

El problema de encontrar la trayectoria óptima Warping es de orden de complejidad exponencial, para reducir éste se utiliza la programación dinámica que disminuye el orden a $O(NM)$ mediante el siguiente algoritmo [43].

Entrada: Matriz de costos $D(n,m) = DTW(X(1:n);Y(1,m))$

*Salida: Trayectoria óptima P^**

La trayectoria óptima $P^=(P_1, \dots, P_L)$ es calculada en orden reverso, por lo tanto se inicia con $P_L(N,M)$.*

En caso $(n,m)=(1,1)$ uno debería tener $l=1$ para finalizar en otro caso:

$$P_{L-1} = \begin{cases} (1, m-1) \text{ si } n = 1 \\ (n-1, 1) \text{ si } m = 1 \\ \text{en otro caso, } \operatorname{argmin}\{D(n-1, m-1), D(n-1, m), D(n, m-1)\} \end{cases}$$

H. Extracción de características en Praat.

Este anexo presenta la forma de obtener las características: duración, intensidad, frecuencia fundamental y las primeras cinco formantes, en el software Praat. Se indica como se obtienen éstas además de presentar la instrucción que permite calcularlas dentro del script. [38, 47]

Duración.

Se calcula dentro de un script mediante la siguiente línea de código:

```
Get total duration
```

Intensidad.

Se calcula en Praat a través de la ventana que se muestra en la figura H.1 donde se obtiene seleccionando el sonido y presionando la opción “To Intensity”, con lo que aparece una ventana donde conservamos el valor por default de “Minimum pitch” mientras que el tamaño de la ventana de análisis se modifica a un valor de 0.001 segundos.

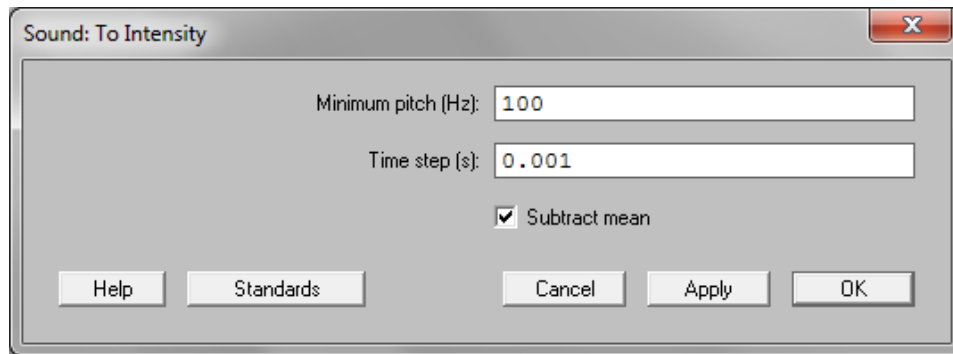


Fig.H.1. Ventana de Praat para calcular la Intensidad de una señal de audio.

La intensidad también se puede calcular a través de las siguientes líneas de código:

```
To Intensity... 100 0.001 yes  
intensidadMedia = Get mean... 0 0 dB
```

Frecuencia Fundamental (F_0).

Para calcular F_0 a través de Praat; primero se selecciona el audio, posteriormente se presiona la opción “Analyse periodicity” que se encuentra del lado derecho de la ventana principal, con lo que aparece un submenú donde se elige el comando “To Pitch” y se despliega otra ventana como la que se muestra en la figura H.2. En ésta se proporcionan los valores para calcular F_0 , de tal forma que se da un avance de 0.001segundos, la mínima frecuencia es 75 Hz y la máxima tiene un valor de 300Hz en voces masculinas mientras que para el resto de voces es de 600Hz.

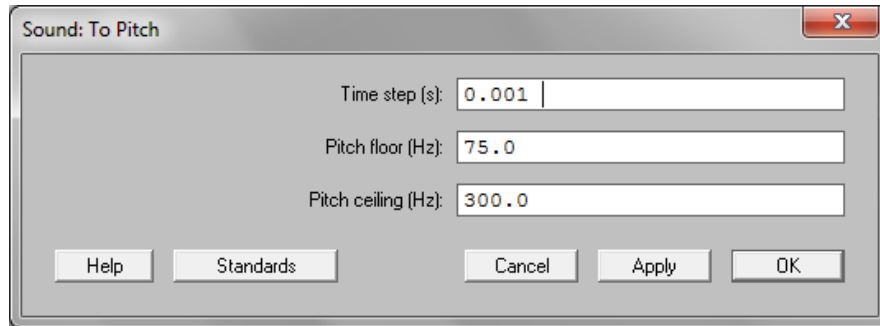


Fig.H.2. Ventana de Praat para calcular el Pitch de una señal de audio.

La ventana de la figura H.2 puede ser reemplazada en un script mediante las siguientes líneas de código:

```
To Pitch... 0.001 75 300  
pitchMedia = Get mean... 0 0 Hertz
```

La primera línea representa los datos que se colocan dentro de la ventana “To Pitch” los cuales aparecen en la misma posición en que se encuentran en ésta. Por otro lado la segunda línea es utilizada para obtener el valor promedio a partir de un recorrido por todos los valores que se tiene para el cálculo de pitch. El valor resultante se mide en Hertz.

Formantes.

Para obtener las formantes a través de Praat primero se selecciona el sonido, posteriormente se elige la opción “Analyse spectrum” que se encuentra de lado derecho de la ventana principal, con lo que aparece un submenú del cual se selecciona el comando “Formant (burg)” y aparece una ventana como la que se encuentra en la figura H.3. En ésta se proporciona el incremento o avance de la ventana de análisis (Time step), la cantidad de formantes a calcular (Max. Number of formants) en nuestro caso el valor es cinco, la frecuencia máxima cuyo valor varía para el caso de una voz femenina adulta es de 5500Hz mientras que para una voz masculina se suele manejar en 5000Hz. Se mantienen tanto Window length y Pre-emphasis con el valor proporcionado por el programa.

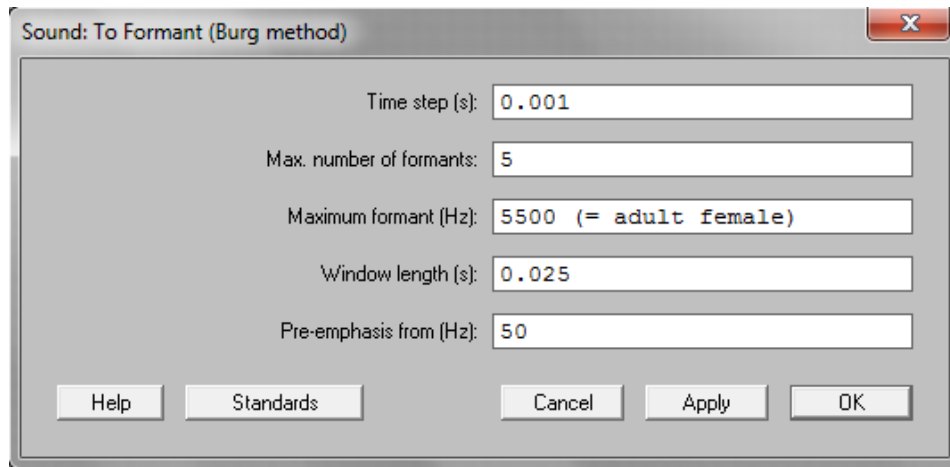


Fig.H.3. Ventana de Praat para calcular las cinco primeras formantes de una señal de audio.

También las formantes se pueden calcular a partir de las siguientes líneas de código:

```
To Formant (burg)... 0.001 5 5500 0.025 50
for i from 1 to 5
    f[i] = Get mean... i 0 0 Hertz
endfor
```

La primer línea se encarga de obtener las formantes y los datos que se proporcionan equivale a los colocados en la ventana de la figura H.3, las líneas que pertenecen al ciclo “for” se usan para obtener el valor promedio de cada una de las formantes.

Los gráficos se puede obtener a través del software Matlab R2012b, donde las primeras dos formantes se grafican utilizando la herramienta “plot” y para cada una de las demás características (pitch, intensidad y duración) se emplea un diagrama de caja con la herramienta “boxplot”.

I. Perceptrones Multicapa (MLP) y Máquinas de Soporte Vectorial (SVM)

Perceptrones Multicapa (MLP)

Las redes neurales son una representación matemática que simula el comportamiento de las neuronas del cerebro. Las redes están compuestas por nodos que se conectan entre ellos mediante conexiones dirigidas y, suponiendo que tenemos dos unidades i y j , la conexión entre ellas sirve para propagar la activación de j a i (a_j) además de que la fuerza y el signo de la conexión se determina a través de un peso numérico $W_{j,i}$. El primer paso para realizar la conexión lo hace i mediante la suma ponderada de sus entradas $in_i = \sum_{j=0}^n W_{j,i} a_j$. Después se aplica la función de activación g sobre la entrada con lo que obtenemos en la salida:

$$a_i = g(in_i) = g\left(\sum_{j=0}^n W_{j,i} a_j\right)$$

En la figura I.1 se encuentra la estructura de un nodo y el comportamiento general de éste. En lo que respecta a la función g ésta puede ser de dos tipos: umbral y sigmoide (logística), además g se diseña para que la unidad se active cuando se proporcionen entradas correctas (cerca de la entrada $a+1$) y cuando las entradas sean erróneas esta se inactiva (cerca a 0).

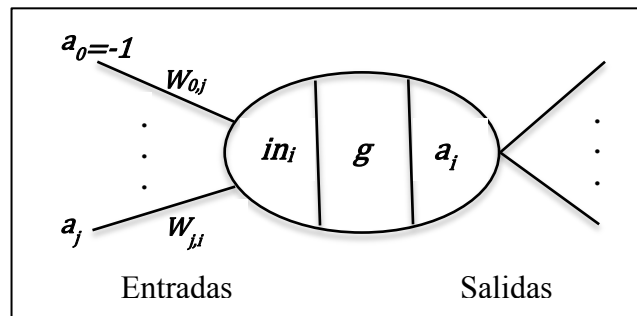


Fig.I.1. Modelo de una neurona que se representa $a_i = g(in_i)$.

Un perceptrón es una red neuronal de una sola capa donde las entradas están conectadas directamente a las salidas, es decir, cada unidad de salida es independiente de otras y no se afectan entre sí. En la figura I.2 podemos observar la estructura de un perceptrón el cual consiste de cinco entradas y tres salidas donde cada unidad de entrada está conectada a su unidad de salida, por lo tanto los pesos de sus conexiones no se afectan entre sí. Las funciones booleanas: **AND**, **OR** y **NOT** pueden ser representadas a través de un perceptrón.

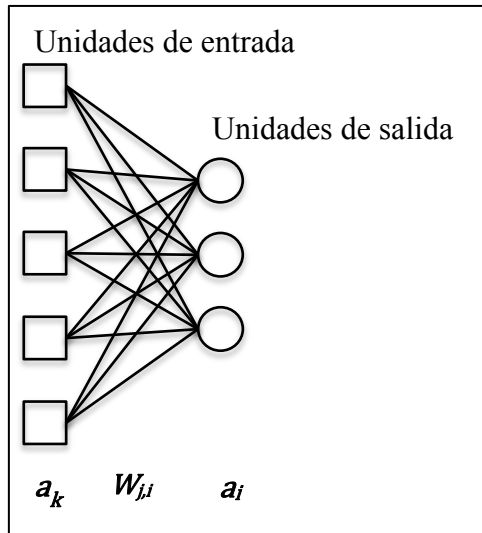


Fig.I.2. Ejemplo de una Red perceptrón.

Un perceptrón multicapa (MLP) es una red con unidades ocultas que ayudan a ampliar el espacio de funciones que se pueden representar mediante la red. Por ejemplo una sola capa oculta (suficientemente grande) puede representar cualquier función continua y con dos capas se pueden representar las funciones discontinuas. El proceso de aprendizaje de la red se realiza mediante un método de retropropagación el cual va modificando los pesos, permitiendo que estos se adapten para representar la función. En la figura I.3 se muestra un ejemplo de un perceptrón multicapa de cinco entradas con una capa oculta de tres unidades.

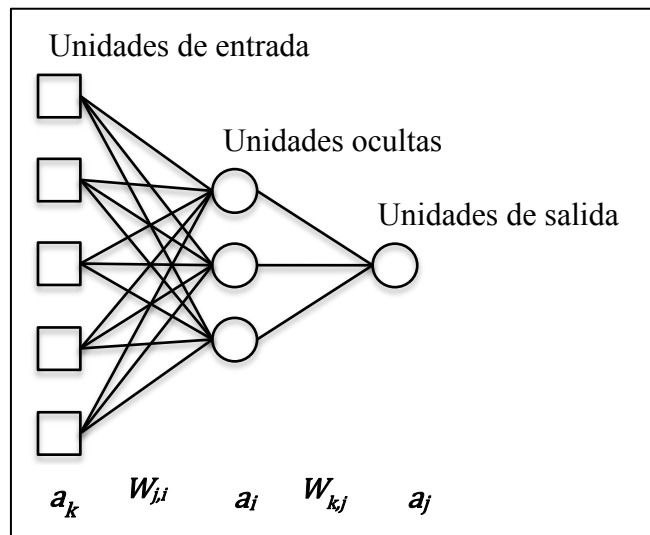


Fig.I.3. Ejemplo de un Perceptrón Multicapa con una capa oculta.

Máquinas de Soporte Vectorial (SVM)

Las Maquinas de Soporte Vectorial (SVM) son también conocidas como Máquinas Núcleo y es un método que implementa un algoritmo de aprendizaje que incluso permite representar funciones no lineales complejas. Parte de un principio donde se tienen dos conjuntos de datos, que están en un mismo plano y no son linealmente separables, donde cada vector de entrada \mathbf{x} corresponde a un nuevo vector de características $F(\mathbf{x})$, definiendo a éste como un conjunto de nuevas funciones que permiten mapear los datos a un nuevo espacio de dimensiones donde los conjuntos si son separables.

El principio como funciona MSV es que dado un conjunto de datos no linealmente separables en un plano, busca la correspondencia de los datos en un nuevo espacio a una dimensión suficientemente alta donde los datos sí son linealmente separables. Pero las MSV deben encontrar el separador lineal óptimo que obtenga el margen más grande entre él y los ejemplos positivos o negativos para cada uno de sus respectivos lados del separador. Para calcular el separador óptimo necesitamos resolver un problema de programación cuadrática que consiste en suponer que inicialmente tenemos ejemplos \mathbf{x}_i con clasificaciones $y_i = \pm 1$, así que el problema consiste en encontrar los valores del parámetro α_i que maximicen la siguiente expresión:

$$\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

Teniendo presente las siguientes restricciones: $\alpha_i \geq 0$ y $\sum_i \alpha_i y_i = 0$.

El separador óptimo encontrado es:

$$h(\mathbf{x}) = \text{signo} \left(\sum_i \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}_i) \right)$$

Los pesos α_i asociados con cada punto de los datos son cero excepto para aquellos puntos más cercanos al separador, lo que denominamos vectores soporte.

La función de núcleo (kernel) generalmente se representa por $K(\mathbf{x}_i, \mathbf{x}_j)$ y es aquella que se aplica a los pares de datos de entrada para evaluar los productos en el espacio de características correspondiente, con el fin de encontrar separadores lineales en este espacio de alta dimensión $F(\mathbf{x})$. Para esto solo basta con remplazar $\mathbf{x}_i \cdot \mathbf{x}_j$ de la anterior sumatoria por una función núcleo $K(\mathbf{x}_i, \mathbf{x}_j)$.

Podemos aplicar diferentes funciones núcleo como pueden ser: el polinomial o la función de base radial (RBF).

- El kernel polinomial está dado por $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{1} + \mathbf{x}_i \cdot \mathbf{x}_j)^d$, que corresponde a un espacio de características cuya dimensión es exponencial en d .
- El kernel de una función de base radial está dado por: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-[\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j]^2}$

J. Palabras utilizadas para determinar el símbolo fonético de Ę.

Para determinar el símbolo fonético SAMPA de la sexta vocal (Ę) se necesita, de un conjunto de sonidos que pertenecen a cada uno de los símbolos dentro del trapecio de vocales IPA. En total se tienen 77 sonidos que fueron segmentados manualmente, los cuales pertenecen a audios que corresponden a palabras encontradas en diferentes transcripciones fonéticas IPA o SAMPA. A continuación se muestra la información de cada símbolo mediante una tabla en la que se indica: el símbolo IPA, su correspondencia SAMPA, la palabra de la cual se segmentó el sonido, la transcripción fonética IPA o SAMPA de la palabra, el idioma en que se pronunció, el género del hablante (masculino o femenino) y el nombre del archivo con el que se almacena.

Símbolo		Palabra	Transcripción Fonética		Idioma	Genero	Archivo
IPA	SAMPA		IPA	SAMPA			
I	i	v _i lle		v i l @	Danés	Masculino	SAMPA1a
		i _d idioma		i D j o m a	Español <España>	Femenina	SAMPA1b
		ch _i ta		tS i t a	Español <México>	Femenina	SAMPA1c
Y	y	n _ü ri		n y r i	Estonio	Masculino	SAMPA2a
		s _ü t		s y t	Hungaro	Masculino	SAMPA2b
		du		d y	Frances	Femenina	SAMPA2c
i	I	r _b i	-	-	Ruso	Masculino	SAMPA3a
		-	gim		Hungaro	Femenina	SAMPA3b
		МЫШЬ	-	m I S	Ruso	Masculino	SAMPA3c
u	}	hus	-	h } : s	Sueco	Masculino	SAMPA4a
		sju	-	-	Sueco	Femenina	SAMPA4b
		sjuk	-	s } : k	Sueco	Femenina	SAMPA4c
W	M	-	Λ:lwn	-	Coreano <Corea del Sur >	Femenina	SAMPA5a
		sju	tw	-	Vietnamita	Femenina	SAMPA5b
		-	-				
U	u	escucha	-	e s k u tS a	Español <España>	Masculina	SAMPA6a
		mucho	-	m u tS o	Español <España>	Femenina	SAMPA6b
		chupan	-	tS u p a n	Español <México>	Femenina	SAMPA6c
I	I	v _i t	-	v I t	Sueco	Masculina	SAMPA7a
		typ	-	t I p	Polaco	Femenina	SAMPA7b
		typ	-	t I p	Polaco	Masculina	SAMPA7c
Y	Y	bytt	-	b Y t	Sueco	Masculina	SAMPA8a
		hübsch	-	h Y p S	Alemán	Masculina	SAMPA8b
		flüsse	f l ə Y s	-	Alemán <Austria>	Femenina	SAMPA8c
o	U	bott	-	b U t	Sueco	Masculina	SAMPA9a
		schutz	-	S U t s	Alemán	Masculina	SAMPA9b

		put	-	p U t	Ingles	Femenina	SAMPA9c
E	e	cero	-	tS e r o	Español <México>	Femenino	SAMPAaAa
		fazer	-	f 6” z e r	Portugués <Portugal>	Femenino	SAMPAAb
		dove	-	d o v e	Italiano	Masculino	SAMPAAc
Ø	2	kolle	-	k 2 l @	Danés	Masculino	SAMPABa
		deux	-	d 2	Francés	Femenino	SAMPABb
		föll	-	f 2 l	Sueco	Masculino	SAMPABc
ə	@/	-	-	-			
		-	-	-			
		-	-	-			
ø	8	buss	-	b u ø s	Sueco	Femenino	SAMPACa
		-	-	-			SAMPACb
		-	-	-			SAMPACc
ɤ	7	sõprus	-	-	Estonio	Femenino	SAMPADa
		kõdu	-	-	Estonio	Masculino	SAMPADb
		uladh		ɤ l̪ u	Irlandes	Femenino	SAMPADc
O	o	cinco	-	T i n k o	Español <México>	Femenino	SAMPAEa
		ocular	-	o k u l a r	Español	Femenino	SAMPAEb
		oda	-	o d a	Español	Femenino	SAMPAEc
ε	E	gesetz	-	g @ z E t s	Alemán	Masculino	SAMPAFa
		gesetz	-	g @ z E t s	Alemán	Femenino	SAMPAFb
		seize	-	s E z	Francés	Masculino	SAMPAFc
Œ	9	neuf	-	n 9 f	Francés	Masculino	SAMPA10a
		peur	p œ ʁ	-	Francés	Femenino	SAMPA10b
		peur	p œ ʁ	-	Francés	Femenino	SAMPA10c
ɜ	3	learn	-	-	Ingles <Americano>	Femenino	SAMPA11a
		urge	-	-	Ingles <Americano>	Masculino	SAMPA11b
		girl	-	-	Ingles <Británico>	Femenino	SAMPA11c
ɛ	3/	-	-	-			
		-	-	-			
		-	-	-			
ʌ	V	cut	-	k V t	Ingles <Americano>	Masculino	SAMPA12a
		cut	-	k V t	Ingles <Británico>	Femenino	SAMPA12b
		struct	-	-	Ingles <Americano>	Masculino	SAMPA12c

ɔ	O	moto	-	m O t o			SAMPA13a
		ontem	-	O n t 6~ j~			SAMPA13b
		trotz	-	t r O t s			SAMPA13c
Æ	{	pat	-	p { t	Ingles	Masculino	SAMPA14a
		bad	-	b { d	Ingles	Femenino	SAMPA14b
		pat	-	p { t	Ingles	Masculino	SAMPA14c
ɐ	6	besser	-	b E s 6	Alemán	Masculino	SAMPA15a
		madeira	-	m 6 “ d 6j r 6	Portugués	Femenino	SAMPA15b
		fazer	-	f 6 “ z e r	Portugués	Femenino	SAMPA15c
A	a	gabriela	-	g a B r j e l a	Español <México>	Femenino	SAMPA16a
		hall	-	h a l	Sueco	Femenino	SAMPA16b
		panna	-	“ p a n n a	Italiano	Femenino	SAMPA16c
œ	&	skörd [57]	-	-	Sueco	Masculino	SAMPA17a
		-	-	-			
		-	-	-			
ɑ	A	father	f a d ə (r)	-	Ingles <Americano>	Femenino	SAMPA18a
		hatt	-	h A t	Noruego	Masculino	SAMPA18b
		pâte	-	p A t	Francés	Masculino	SAMPA18c
ɒ	Q	pot	-	p Q t	Ingles <Británico>	Femenino	SAMPA19a
		lot	-	-	Ingles <Británico>	Masculino	SAMPA19b
		bother	b D t ə	-	Ingles <Británico>	Masculino	SAMPA19c
ə	@	bitte	-	b I t @	Alemán	Masculino	SAMPA1Aa
		allow	-	a l a U	Ingles <Americano>	Masculino	SAMPA1Ab
		felizes	-	f @ l i z @ S	Portugués <Brasil>	Femenino	SAMPA1Ac

K. Determinar Fonemas.

El Anexo K contiene los resultados que pertenecen al paso de la metodología “Determinar Fonemas”, donde los resultados de cada una de las pruebas se colocan en una tabla como la que se muestra en la figura K.1, en la que se indica la palabra antes y después de realizar el análisis de la vocal, y la cual es remarcada con negritas; en lo que respecta a las consonantes permanecerán constantes.

No.	Palabra		Palabra Final
	Español	Zapoteco	
1			

Fig.K.1. Ejemplo de tabla de final de las palabras analizadas.

Las vocales que presentan incertidumbre son cinco: A, E, I, O y U, en cualquiera de las tres realizaciones (cortadas, quebradas y sencillas); en lo que respecta a la sexta vocal Ë como no se ocupó dentro de las palabras que tienen conflicto, sólo se utilizará dentro de los audios prototipo al realizar el análisis de la vocal E debido a la semejanza acústica entre ambas. A continuación presentamos cada uno de estos los resultados para las vocales.

Vocal A.

La cantidad inicial de fonemas conocidos para la vocal A son 101, los cuales se encuentran repartidos en sus tres realizaciones como lo ilustra la tabla K.1.

Tabla.K.1. Cantidad de fonemas de la vocal A en sus tres diferentes realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
A	80
A'	17
A'A	4

En la tabla K.2 se encuentran las palabras de donde se tomaron los 15 audios prototipo de la vocal A, donde cada realización de la vocal cuenta con la misma representación. Además se indica el número de repetición de la palabra de la cual fueron tomados estos sonidos.

Tabla.K.2. Palabras de donde se extrajeron los audios prototipo para el análisis DTW de la vocal A.

No.	Español	Zapoteco	Repetición
1	BOCA	RA'	1
2	AHÍ	LENA'	1
3	CABELLO	GUITZA'	1
4	OREJA	NAGA'	1
5	QUE	BIDA'N	1
6	AMARILLO	GACHI	1
7	ÁRBOL	YAGA YULA	1

8	AGUA	NISA	1
9	ALA	XILA	1
10	TRES	TZONAN	1
11	BEBER	RA'A	1
12	BEBER	RA'A	2
13	MANO	NA'A	1
14	PARTIR	LA'AN	1
15	VERDE	YA'A	1

Los audios de prueba para la vocal A provienen de dieciocho palabras que tienen una ortografía que no pudo ser corroborada en algún diccionario, las cuales se muestran en la tabla K.3. La vocal analizada se resalta en negritas para su mejor visualización, el elemento <I> indica que al analizar la tabla de costos no se pudo determinar un tipo de realización para la vocal así que es mejor eliminarla de la etapa de análisis. Como observación final de esta tabla se debe resaltar que los diptongos que contengan la vocal A no se incluyen en el proceso de comparación, un ejemplo de esto ocurre con la palabra RAGUABA en la que solo se analizaron las vocales del inicio y del final de la palabra.

Tabla.K.3. Fonemas finales para las tres realizaciones de la vocal A.

No.	Palabra inicial		Palabra Final
	Español	Zapoteco	
1	ALGUNOS	NULAN NA'	NULA'AN NA'
2	APRETAR	QUICHAN	QUICHA'AN
3	APUÑALAR	QUIBA	QUIBA'
4	ARAÑAR	INIZA	INIZA'
5	ATAR	SCHOGAN	SCHOGA'AN
6	CALIDO	SZLA	SLZA
7	CAZAR	RAGUABA	RAGUABA
8	GOLPEAR	IZANA	IZANA'A
9	GORDO	MALLÍ	MALLÍ
10	JALAR	ZAN	ZA'N
11	LAVAR	GUIVA	GUIV<I>
12	MORDER	RIÑA'	RIÑA'
13	NIEBLA	LANZA	LANZA
14	NUBE	CHULA	CHULA'
15	REÍR	RIYA'	RIYA'
16	RODILLA	RARRIBA	RARRIB<I>
17	SOPLAR	LUBA	LUB<I>
18	VOMITAR	GUEBA	GUEB<I>

Para terminar el análisis de los fonemas indeterminados de la vocal A falta analizar a dos palabras que utilizan el diptongo AA para lo cual a los quince audios prototipo se les agregaron dos ejemplos más que pertenecen a las repeticiones de la palabra "GUIEBAA" (CIELO).

Tabla.K.4. Ejemplos del fonema AA que pertenece a la palabra GUIEBAA (CIELO) en su primera y segunda repetición.

No.	Español	Zapoteco	Repetición
1	CIELO	GUIEBAA	1
2	CIELO	GUIEBAA	2

En la tabla K.5 se encuentran las palabras desconocidas que utilizan el diptongo AA, además también se muestra la palabra después de determinar el tipo de sonido de la vocal A.

Tabla.K.5. Fonemas finales para el diptongo AA, los cuales se encuentran resaltados dentro de las palabras que lo integran.

No.	Palabra inicial		Palabra Final
	Español	Zapoteco	
1	VOLAR	SAA	SA'
2	CUERNO	LUZUBAA	LUZUB<I>

Al inicio del análisis la vocal A contaba con 25 sonidos que estaban en un estado indeterminado (I), después de realizar éste sólo el 20% de éstos continúan en un estado desconocido como se muestra en la tabla K.6.

Tabla.K.6. Tabla de cambios de la vocal A. Indica cuales fueron los cambios que sufrieron los grafemas de la vocal A al realizar el análisis para determinar los fonemas.

Cambio	No. Grafemas	Porcentaje [%]
A' → A'	3	12
A → A'A	4	16
A → A	8	32
A → A'	4	16
AA → A'	1	4
A → I	5	20
Total	25	100%

En la tabla K.7 se muestra un resumen de los cambios que sufre la base de datos en el número de audios de la vocal A en sus tres tipos de realizaciones. Se debe tener presente que entre las palabras conocidas se encuentran dos que tienen una sola repetición y que contienen a las vocales A y A' lo que causa que el número de sonidos finales para éstas vocales no aumente al doble.

Tabla.K.7. Esta tabla muestra cómo se vieron afectados en cantidad los audios en las diferentes realizaciones de la vocal.

Tipo de Realización	Una repetición			Dos repeticiones
	Audios Nuevos	Audios Iniciales	Audios Finales	Audios Totales
A	8	80	88	175
A'	8	17	25	49
A'A	4	4	8	16

Vocal E.

La base de datos contiene 29 palabras que contienen algún fonema de la vocal E en cualquiera de sus tres realizaciones; distribuidos como se muestra en la tabla K.8.

Tabla.K.8. Cantidad de Fonemas de la vocal E en sus tres realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
E	27
E'	3
E'E	3

Debido a que semejanza acústica de las vocales E y Ë, se necesita utilizar los audios de esta última en sus tres tipos de realizaciones los cuales se encuentran distribuidos como se muestra en la tabla K.9.

Tabla.K.9. Cantidad de Fonemas de la vocal Ë en sus tres realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
Ë	2
Ë'	6
Ë'Ë	1

Los audios prototipo pertenecen tanto a las vocales E y Ë en sus tres realizaciones. Los sonidos que se utilizaron pertenecen a las palabras que se encuentran en la tabla K.10.

Tabla.K.10. Palabras de donde se extrajeron los audios prototipo de la vocal Ę y E en sus tres realizaciones, para el análisis DTW.

No.	Español	Zapoteco	Repetición
1	EL	LĚUBI'	1
2	EL	LĚUBI'	2
3	ESPOSO	BĚ'M BYU QUIA	1
4	HOMBRE	BĚ'M BYU	1
5	POCO	LĚ'TIN	1
6	TU	LĚ'	1
7	YO	NĚ'DI'	1
8	CARNE	BĚ'ĚLA	1
9	CARNE	BĚ'ĚLA	2
10	SEMILLA	LE'XI	1
11	SEMILLA	LE'XI	2
12	MUCHOS	BACASIDE'	1
13	MUCHOS	BACASIDE'	2
14	BARRIGA	LE'E	1
15	BARRIGA	LE'E	2
16	CAMINO	NEZA	1
17	DIENTE	LEYA'	1
18	ESPALDA	EXCU'ZU'	1
19	ESTRELLA	BELO	1
20	GIRAR	CHECHOGA	1

En la tabla K.11 se encuentran las siete palabras desconocidas antes y después de determinar el fonema de la vocal E.

Tabla.K.11. Fonemas finales para las tres realizaciones de las vocales E y Ę.

No.	Palabra inicio		Palabra Final
	Español	Zapoteco	
1	ACHRSE	CUEGE	CUEGE'
2	CAER	GUEZO	GUĚ'ZO
3	CUELLO	GUERU	GUERU
4	LAGO	ELGUN	Ě'LGUN
5	NOSOTROS	NETURU	NĚ'ĚTURU
6	RESPIRAR	REZIABE	RĚ'ĚZIABE'
7	VOMITAR	GUEBA	GUĚ'ĚBA

El número de cambios que sufrieron los sonidos indeterminados de E se muestra en la tabla K.12 en donde ningún sonido continuo siendo desconocido.

Tabla.K.12. Tabla de cambios de la vocal E. Indica cuales fueron los cambios que sufrieron los grafemas de la vocal E al realizar el analisis para determinar los fonemas.

Cambio	No. Grafemas	Porcentaje [%]
E → E	1	12.50
E → E'	2	25.00
E → Ë'	2	25.00
E → Ë'Ë	3	37.50
Total	8	100%

En la tabla K.13 se muestra el número de audios finales que pasaran a la etapa de análisis.

Tabla.K.13. Esta tabla muestra cómo se vieron afectados en cantidad los audios en las diferentes realizaciones de las vocales.

Tipo de Realización	Una repetición			Dos repeticiones
	Audios Nuevos	Audios Iniciales	Audios Finales	Audios Totales
E	1	27	28	55
E'	2	3	5	10
E'E	0	3	3	6
Ë	0	2	2	4
Ë'	2	6	8	16
Ë'Ë	3	1	4	8

Vocal I.

En la base de datos existen 45 palabras que contienen al menos un fonema de la vocal I en cualquiera de sus tres realizaciones y se encuentran distribuidos como se muestra en la tabla K.14.

Tabla.K.14. Cantidad de Fonemas de la vocal I en sus tres realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
I	36
I'	11
I'I	2

De los 49 fonemas que se encuentran en la tabla K.14 se seleccionan los audios prototipos que pertenecen a las palabras que se encuentran en la tabla K.15.

Tabla.K.15. Palabras de donde se extrajeron los audios prototipo de la vocal I en sus tres realizaciones, para el análisis DTW.

No.	Español	Zapoteco	Repetición
1	AMARILLO	GACHI	1
2	AÑO	GUIZA	1
3	ELLOS	LEUYACABI	1
4	ESE	BINA'	1
5	ESTE	DANI	1
6	PIOJO	BECHI'	1
7	PIOJO	BECHI'	2
8	YO	NĒ'DI'	1
9	YO	NĒ'DI'	2
10	NIÑO	XCUIDI'	1
11	PESADO	ZI'I	1
12	PESADO	ZI'I	2
13	RECTO	LI'I	1
14	RECTO	LI'I	2

Los audios de prueba corresponden a las 15 palabras cuya ortografía se considera dudosa como se muestran en la tabla K.16, resaltando en negritas el sonido de la vocal analizada.

Tabla.K.16. Fonemas finales para las tres realizaciones de la vocal I.

No.	Palabra Inicio		Palabra Final
	Español	Zapoteco	
1	APRETAR	QUICHAN	QUI'CHAN
2	APUÑALAR	QUIBA	QUIBA
3	ARAÑAR	INIZA	I'NI'ZA
4	BLANCO	CHIGUICHI	CHIGUICHI'
5	GOLPEAR	IZANA	I'ZANA
6	GORDO	MALLÍ	MALLI'
7	HIGADO	GUIRRI	GUIRRI
8	HINCHARSE	GUI	GUI'
9	LABAR	GUIVA	GUIVA
10	LISO	LILI	LILI'
11	MONTAÑA	GUIZIDAU	GUI'ZI'DAU
12	MORDER	RIÑA'	RI'ÑA
13	REIR	RIYA'	RI'YA'
14	RODILLA	RARRIBA	RARRI'BA
15	SUCIO	GUCHI	GUCHI

De las 21 vocales indeterminadas el 61.9% de ellas cambio a una realización cortada como se muestra en la tabla K.17.

Tabla.K.17. Tabla de cambios de la vocal I. Indica cuales fueron los cambios que sufrieron los grafemas de la vocal I al realizar el analisis para determinar los fonemas.

Cambio	No. Grafemas	Porcentaje [%]
I → I	8	38.095
I → I'	13	61.905
Total	21	100%

En la tabla K.18 se muestra la actualización en el número de audios de la vocal I que pasan a la etapa de análisis.

Tabla.K.18. Esta tabla muestra cómo se vieron afectados en cantidad los audios en las diferentes realizaciones de las vocales.

Tipo de Realización	Una repetición			Dos repeticiones
	Audios Nuevos	Audios Iniciales	Audios Finales	Audios Totales
I	8	36	44	88
I'	13	11	24	48
I'I	0	2	2	4

Vocal O.

La vocal O se utiliza en 12 palabras que se encuentran distribuidas como se indica en la siguiente tabla K.19.

Tabla.K.19. Cantidad de Fonemas de la vocal O en sus tres realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
O	11
O'	0
O'O	2

La vocal O no cuenta con ejemplos para la realización O', por lo que las comparaciones sólo se efectuaron entre la realización sencilla y quebrada. Los audios prototipos fueron segmentados de 9 palabras que se encuentran en la tabla K.20, donde se indica con negritas la vocal utilizada.

Tabla.K.20. Palabras de donde se extrajeron los audios prototipo de la vocal O en sus tres realizaciones, para el análisis DTW.

No.	Español	Zapoteco	Repetición
1	NO	O'O	1
2	NO	O'O	2
3	FLOR	YO'O	1
4	FLOR	YO'O	2
5	CABEZA	GUICHORO	1
6	CORTAR	CHOGA	1
7	ESTRELLA	BELO	1
8	PIEDRA	YO	1
9	TRES	TZONAN	1

En la tabla K.21 se encuentran las palabras de las que fueron segmentados los audios de prueba, donde se resalta que se analizó el diptongo OO que pertenece a la palabra ROO. Además también se indica la ortografía final de éstas.

Tabla.K.21. Fonemas finales para las tres realizaciones de la vocal O.

No.	Palabra		Palabra Final
	Español	Zapoteco	
1	ATAR	SCHOGAN	SCHOGAN
2	ARENA	ROO	RO
2	CAER	GUEZO	GUEZ<I>

De las 6 vocales indeterminadas el 33% continúan en este estado como se muestra en la tabla K.22.

Tabla.K.22. Tabla de cambios de la vocal O. Indica cuales fueron los cambios que sufrieron los grafemas de la vocal O al realizar el analisis para determinar los fonemas.

Cambio	No. Grafemas	Porcentaje [%]
O → O	1	33.33
OO → O	1	33.33
OO → <I>	1	33.33
Total	3	100%

En la tabla K.23 se muestra los audios totales después del análisis de la vocal O.

Tabla.K.23. Esta tabla muestra cómo se vieron afectados en cantidad los audios en las diferentes realizaciones de las vocales.

Tipo de Realización	Una repetición			Dos repeticiones
	Audios Nuevos	Audios Iniciales	Audios Finales	Audios Totales
O	2	11	13	26
O'	0	0	0	0
O'O	0	2	2	4

Vocal U.

El total de palabras que contienen al menos algún tipo de realización de la vocal U es 36 y se encuentran distribuidos como se indica la tabla K.24.

Tabla.K.24. Cantidad de Fonemas de la vocal U en sus tres realizaciones.

Tipo de Realización	Audios Iniciales (en una Repetición)
U	30
U'	9
U'U	0

Los audios prototipo fueron segmentados de las 10 palabras que se encuentran en la tabla K.25, donde las vocales utilizadas se encuentran resaltadas con negritas.

Tabla.K.25. Palabras de donde se extrajeron los audios prototipo de la vocal U en sus tres realizaciones, para el análisis DTW.

No.	Español	Zapoteco	Repetición
1	MUJER	GULA	1
2	CUERDA	DU	1
3	DELGADO	LACU	1
4	DOS	CHUPAN	1
5	ESPOSA	GULA QUIA	1
6	ESPALDA	EXCU'ZU'	1
7	QUIEN	NU'	1
8	PLUMA	DU'BI'	1
9	CINCO	GA'YU'N	1
10	PERRO	BECU'	1

En la tabla K.26 se encuentran las palabras desconocidas que utilizan la vocal U antes y después de realizar el análisis.

Tabla.K.26. Fonemas finales para las tres realizaciones de la vocal U.

No.	Palabra		Palabra Final
	Español	Zapoteco	
1	ALGUNOS	NULAN NA'	NULAN NA
2	CHUPAR	ZUPIA	ZUPIA
3	CUELLO	GUERU	GUERU'
4	CUERNO	LUZUBAA	LUZUBAA
5	LAGO	ELGUN	ELGUN
6	NOSOTROS	NETURU	NETUR<I>
7	NUBE	CHULA	CHULA
8	OLER	NUPIAN	N<I>PIAN
9	PIERNA	ZUN	Z<I>N
10	SOPLAR	LUBA	LUBA
11	SUCIO	GUCHI	GUCHI

En la tabla K.27 se muestra que 23.1% de las vocales analizadas continúan siendo indeterminadas.

Tabla.K.27. Tabla de cambios de la vocal U. Indica cuales fueron los cambios que sufrieron los grafemas de la vocal U al realizar el analisis para determinar los fonemas.

Cambio	No. Grafemas	Porcentaje [%]
U → U	9	69.231
U → U'	1	7.692
U → <I>	3	23.077
Total	13	100%

Los audios totales que pasan a la etapa de análisis aumentan a 78 para la vocal U y a 20 para la U'U cortada como se encuentra especificado en la tabla K.28.

Tabla.K.28. Esta tabla muestra cómo se vieron afectados en cantidad los audios en las diferentes realizaciones de las vocales.

Tipo de Realización	Una repetición			Dos repeticiones
	Audios Nuevos	Audios Iniciales	Audios Finales	Audios Totales
U	9	30	39	78
U'	1	9	10	20
U'U	0	0	0	0

L. Resultados de los Métodos de Reconocimiento de Patrones.

Este anexo presenta los resultados de las 48 pruebas realizadas con los clasificadores MLP y MSV (este último implementa dos kernel: función de base radial y polinomial) utilizando como atributos la variación de las cinco primeras formantes que se extrajeron de ambas bases de datos. Las clases que se identifican son seis cuando se toma en cuenta los tipos de vocal (A, E, È, I, O, U) y dieciocho para el conjunto total de vocales del zapoteco. El resultado se muestra a través de dos tablas:

1. La tabla de Datos indica: la procedencia de los datos (Oaxaca o Veracruz), el tipo de clasificador y la cantidad de vocales a clasificar (seis o dieciocho)
2. La tabla de Resultados indica: el porcentaje de las instancias que fueron correctamente clasificadas.

La sintaxis con que se indica la cantidad de atributos utilizados en una prueba es F1-X donde: 1 representa la primera formante y X señala la última formante empleada en la prueba su valor puede variar desde dos hasta cinco.

A continuación se presentan los resultados de las pruebas realizadas. En amarillo sobresale el mejor resultado mientras que en rojo se indica el peor resultado.

Base de datos de la variante de Choapan del estado de Veracruz.

Tabla.L.1. Datos de la primera serie de pruebas para el clasificador MSV.

Datos	
Estado	Veracruz
Clasificador	MSV
Clases	A, E, È, I, O, U

Tabla. L.2. Resultados de la primera serie de pruebas con el clasificador MSV.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MSV: poly F1-5	86.90
MSV: poly F1-4	87.88
MSV: poly F1-3	87.88
MSV: poly F1-2	86.90
MSV: RBF F1-5	60.22
MSV: RBF F1-4	60.06
MSV: RBF F1-3	60.22
MSV: RBF F1-2	50.24

Tabla.L.3. Datos de la primera serie de pruebas para el clasificador MLP.

Datos de la Prueba	
Estado	Veracruz
Clasificador	MLP
Clases	A, E, Ë, I, O, U

Tabla. L.4. Resultados de la primera serie de pruebas con el clasificador MLP.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MLP F1-5	90.34
MLP F1-4	91.98
MLP F1-3	91.32
MLP F1-2	91.98

Tabla.L.5. Datos de la segunda serie de pruebas para el clasificador MSV.

Datos de la Prueba	
Estado	Veracruz
Clasificador	MSV
Clases	A, A', A'A, E, E', E'E, Ë, Ë', Ë'Ë, I, I', I'I, O, O', O'O, U, U', U'U

Tabla. L.6. Resultados de la segunda serie de pruebas con el clasificador MSV.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MSV poly F1-5	62.52
MSV poly F1-4	63.33
MSV poly F1-3	63.33
MSV: poly F1-2	62.84
MSV: RBF F1-5	28.47
MSV: RBF F1-4	28.64
MSV: RBF F1-3	28.47
MSV: RBF F1-2	28.31

Tabla.L.7. Datos de la segunda serie de pruebas para el clasificador MLP.

Datos de la Prueba	
Estado	Veracruz
Clasificador	MLP
Clases	A, A', A'A, E, E', E'E, Ë, Ë', Ë'Ë, I, I', I'I, O, O', O'O, U, U', U'U

Tabla.L.8. Resultados de la segunda serie de pruebas con el clasificador MLP.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MLP F1-5	65.13
MLP F1-4	66.12
MLP F1-3	66.77
MLP F1-2	67.26

Base de datos de la variante del distrito de Choapan.

Tabla.L.9. Datos de la primera serie de pruebas para el clasificador MSV.

Datos de la Prueba	
Estado	Oaxaca
Clasificador	MSV
Clases	A, E, Ë, I, O, U

Tabla.L.10. Resultados de la primera serie de pruebas con el clasificador MSV.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MSV: poly F1-5	74.8187
MSV: poly F1-4	75.1986
MSV: poly F1-3	73.0397
MSV: poly F1-2	70.4318
MSV: RBF F1-5	55.2677
MSV: RBF F1-4	56.9430
MSV: RBF F1-3	54.3869
MSV: RBF F1-2	36.9430

Tabla.L.11. Datos de la primera serie de pruebas para el clasificador MLP.

Datos de la Prueba	
Estado	Oaxaca
Clasificador	MLP
Clases	A, E, Ë, I, O, U

Tabla. L.12. Resultados de la primera serie de pruebas con el clasificador MLP.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MLP F1-5	77.3575
MLP F1-4	77.5475
MLP F1-3	75.1468
MLP F1-2	70.8463

Tabla.L.13. Datos de la segunda serie de pruebas para el clasificador MSV.

Datos de la Prueba	
Estado	Oaxaca
Clasificador	MSV
Clases	A, A', A'A, E, E', E'E, Ë, Ë', Ë'Ë, I, I', I'I, O, O', O'O, U, U', U'U

Tabla. L.14. Resultados de la segunda serie de pruebas con el clasificador MSV.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MSV poly F1-5	59.0328
MSV poly F1-4	58.6874
MSV poly F1-3	56.5285
MSV: poly F1-2	53.1261
MSV: RBF F1-5	34.8359
MSV: RBF F1-4	34.6632
MSV: RBF F1-3	31.1917
MSV: RBF F1-2	25.3195

Tabla.L.15. Datos de la segunda serie de pruebas para el clasificador MLP.

Datos de la Prueba	
Estado	Oaxaca
Clasificador	MLP
Clases	A, A', A'A, E, E', E'E, Ë, Ë', Ë'Ë, I, I', I'I, O, O', O'O, U, U', U'U

Tabla. L.16. Resultados de la segunda serie de pruebas con el clasificador MLP.

Resultados	
Método	Instancias Correctamente Clasificadas [%]
MLP F1-5	61.3472
MLP F1-4	60.9672
MLP F1-3	59.6373
MLP F1-2	54.6805

M. Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě).

En general todos los experimentos realizados para el paso de la metodología “Determinación del Símbolo Fonético SAMPA de la Sexta Vocal (Ě)” suman 24 utilizando los software: Matlab y Praat. Los resultados se presentan en una tabla donde se indica: el símbolo IPA con su equivalencia en SAMPA, la palabra de la cual fue segmentado este símbolo, el idioma en que se pronunció la palabra, el género del hablante (femenino o masculino), la cantidad de audios de prueba que se asemejan al sonido de este símbolo, el porcentaje de éstos con respecto al total de audios de prueba y el promedio de los costos DTW para la cantidad de audios que coinciden con este símbolo.

Para la base de datos de la variante de Choapan del estado de Veracruz los audios de prueba provienen de las palabras que se muestran en la tabla M.1, donde la letra resaltada en negritas indica el sonido de la vocal que se utilizó.

Tabla.M.1. Palabras de donde provienen los sonidos empleados en el análisis para la determinación del símbolo fonético.

Vocal	Palabras
Ě*	LĚUBI', LĚE'
Ě'	Ě'LGUN, BĚ'M BYU, LĚ'TIN, LĚ', NĚ'DI, BENĚ', BĚ'M BYU QUIA
Ě'Ě	BĚ'ĚLA, NĚ'ĚTUR<I>, RĚ'ĚZIABE', GUĚ'ĚB<I>
E	EXCU'ZU', LENA', NEZA, DE, GUERU', TEZA, LEYA', BELO, CHECHOGA, ĤE, BELA, BENĚ', DZEN, GUEGALA, BEN XHNI'A, PENSARE, BECU', BECHI', CHĤEBA, NEZIU, ZEDE', REN, BELA, GUELA, BE, GUI'BELA
E'	LE'XI, CUEGE', ZEDE', RĚ'ĚZIABE', BACASIDE'
E'E	LE'E, RE'EN, LE'EDA

*Debido a la falta de audios para la vocal Ě se utilizó los sonidos segmentados de los diptongos: eu y ěe'.

Resultados del método DTW utilizando la herramienta Matlab.

Resultados para la variante del distrito de Choapan.

Se realizan en total seis pruebas las cuales consisten en analizar cada una de las realizaciones de las vocales Ę y E.

Prueba I. DTW aplicado a la vocal Ę en su realización cortada (Ę').

Tabla.M.2 Resultados DTW para la vocal Ę'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	Buss	Sueco	Femenino	18	60.000	7.695
ʌ	V	cut	Ingles	Masculino	4	13.333	10.174
ɔ	O	ontem	Portugués	Femenino	2	6.667	9.429
ɐ	6	madeira	Portugués	Femenino	2	6.667	9.175
ʊ	U	Book	Ingles	Femenino	1	3.333	9.231
ø	8	Full	Sueco	Masculino	1	3.333	9.615
ɑ	A	Father	Ingles	Femenino	1	3.333	10.963
o	O	Ocular	Español	Femenino	1	3.333	10.775
					30	100 %	

Prueba II. DTW aplicado a la vocal Ę en su realización quebrada (Ę'Ę).

Tabla.M.3 Resultados DTW para la vocal Ę'Ę.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	21	70.000	9.490
ɯ	}	sjuk	Sueco	Femenino	5	16.667	12.125
Y	Y	syll	Sueco	Masculino	2	6.667	15.183
ɐ	6	madeira	Portugués	Femenino	1	3.333	11.446
ø	2	föll	Sueco	Masculino	1	3.333	15.237
					30	100 %	

Prueba III. DTW aplicado a la vocal Ę en su realización sencilla (Ę).

Tabla.M.4. Resultados DTW para la vocal Ę.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	30	100	8.349
					30	100 %	

Prueba IV. DTW aplicado a la vocal e en su realización cortada (E').

Tabla.M.5. Resultados DTW para la vocal E'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɔ	O	ontem	Portugués	Femenino	14	56.000	8.430
ə	8	buss	Sueco	Femenino	4	16.000	8.894
ʊ	U	book	Ingles	Femenino	2	8.000	9.880
ʌ	V	cut	Ingles	Femenino	1	4.000	14.411
ɐ	6	madeira	Portugués	Femenino	1	4.000	10.091
ə	8	full	Sueco	Masculino	1	4.000	10.293
a	a	panna	Italiano	Masculino	1	4.000	12.872
Y	Y	syll	Sueco	Masculino	1	4.000	13.930
					25	100 %	

Prueba V. DTW aplicado a la vocal e en su realización quebrada (E'E).

Tabla.M.6. Resultados DTW para la vocal E'E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ə	8	buss	Sueco	Femenino	13	43.333	8.632
Y	Y	flüsse	Alemán	Femenino	3	10.000	11.161
ʊ	U	book	Ingles	Femenino	3	10.000	10.086
ɐ	}	sjuk	Sueco	Femenino	2	6.667	10.564
ə	8	full	Sueco	Masculino	2	6.667	9.616
ɔ	O	ontem	Portugués	Femenino	2	6.667	10.312
ɑ	A	father	Ingles	Femenino	2	6.667	12.719
Y	Y	syll	Sueco	Masculino	2	6.667	11.606
ɐ	6	madeira	Portugués	Femenino	1	3.333	9.896
					30	100 %	

Prueba VI. DTW aplicado a la vocal e en su realización sencilla (E).

Tabla.M.7. Resultados DTW para la vocal E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ə	8	buss	Sueco	Femenino	27	90.000	8.586
ə	8	full	Sueco	Masculino	1	3.333	11.592
u	u	escucha	Español	Masculino	1	3.333	19.924
ʊ	U	book	Ingles	Femenino	1	3.333	13.587
					30	100 %	

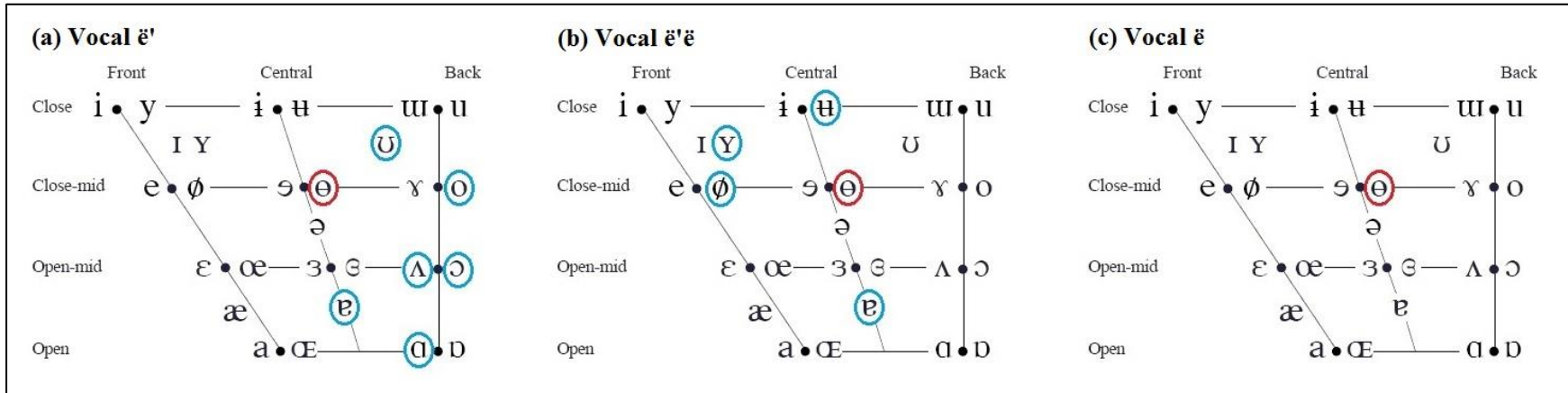


Fig.M.1. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal \tilde{E} en sus tres realizaciones para audios de prueba que pertenecen a la variante del distrito de Choapan. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

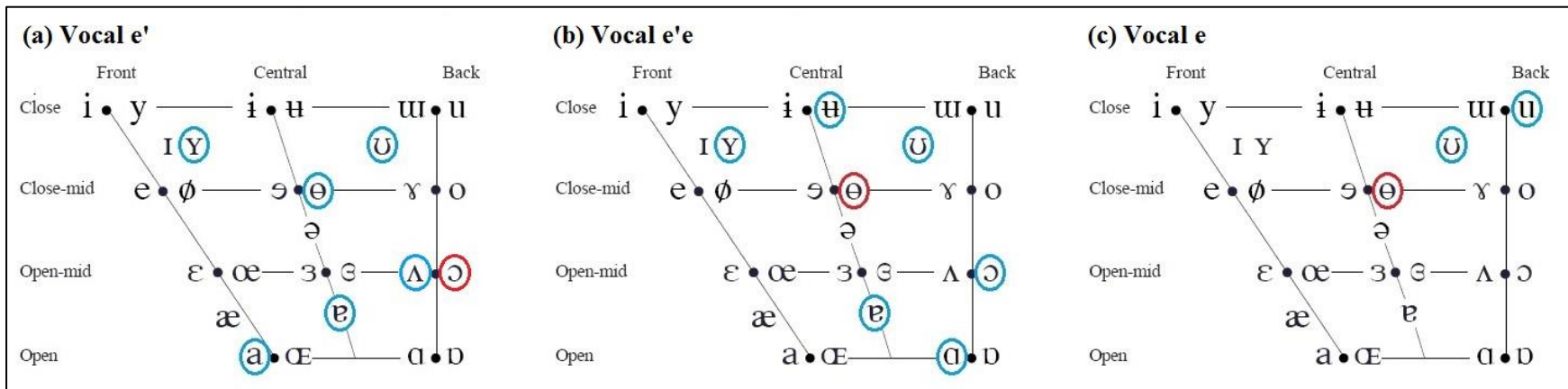


Fig.M.2. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal E en sus tres realizaciones para audios de prueba que pertenecen a la variante del distrito de Choapan. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

Resultados para la variante de Choapan del estado de Veracruz.

Se realizan en total seis pruebas las cuales consisten en analizar cada una de las realizaciones de las vocales \ddot{E} y E.

Prueba VII. DTW aplicado a la vocal \ddot{e} en su realización cortada (\ddot{E}').

Tabla.M.8. Resultados DTW para la vocal \ddot{E}' .

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	10	71.428	16.323
ʊ	U	book	Ingles	Femenino	2	14.286	19.707
ʌ	V	cut	Ingles	Masculino	1	7.143	20.387
e	e	test	Polaco	Masculino	1	7.143	21.386
					14	100 %	

Prueba VIII. DTW aplicado a la vocal \ddot{e} en su realización quebrada ($\ddot{E}'\ddot{E}$).

Tabla.M.9. Resultados DTW para la vocal $\ddot{E}'\ddot{E}$.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	7	87.500	15.375
ʊ	U	book	Ingles	Femenino	1	12.500	20.174
					8	100 %	

Prueba IX. DTW aplicado a la vocal \ddot{e} en su realización sencilla (\ddot{E}).

Tabla.M.10. Resultados DTW para la vocal \ddot{E} .

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	4	100	8.349
					4	100 %	

Prueba X. DTW aplicado a la vocal e en su realización cortada (E').

Tabla.M.11. Resultados DTW para la vocal E'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	7	70.000	14.808
ʌ	V	cut	Ingles	Femenino	3	30.000	17.512
					10	100 %	

Prueba XI. DTW aplicado a la vocal e en su realización quebrada (E'E).

Tabla.M.12. Resultados DTW para la vocal E'E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	5	83.333	17.187
E	e	test	Polaco	Masculino	1	16.667	20.257
					6	100 %	

Prueba XII. DTW aplicado a la vocal e en su realización sencilla (E).

Tabla.M.13. Resultados DTW para la vocal E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ø	8	buss	Sueco	Femenino	23	76.667	15.133
a	a	panna	Italiano	Masculino	2	6.667	17.546
ʌ	V	cut	Ingles	Femenino	2	6.667	19.444
ɒ	Q	lot	Ingles	Masculino	1	3.333	26.492
ʊ	U	book	Ingles	Femenino	1	3.333	18.117
ɔ	O	ontem	Portugués	Femenino	1	3.333	20.056
					30	100 %	

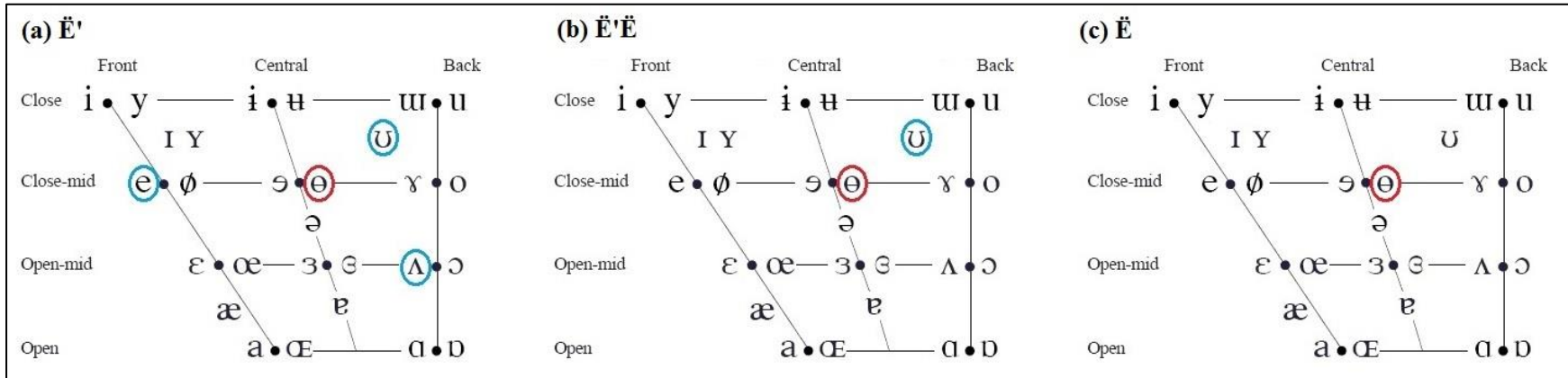


Fig.M.3. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal Ë en sus tres realizaciones para audios de prueba que pertenecen a la variante de Choapan del estado de Veracruz. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

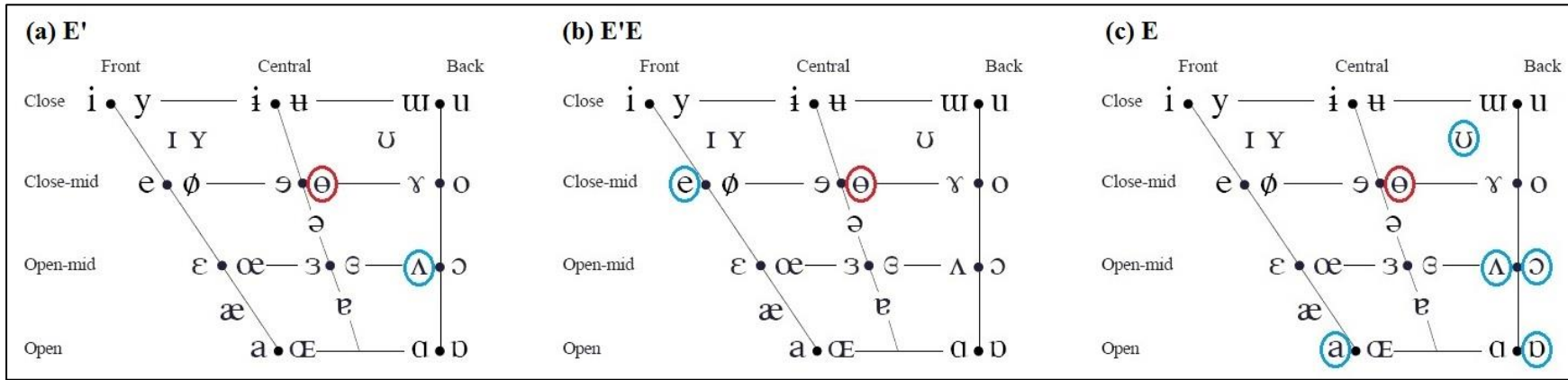


Fig.M.4. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal E en sus tres realizaciones para audios de prueba que pertenecen a la variante de Choapan del estado de Veracruz. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

Resultados del método DTW utilizando la herramienta Praat.

Resultados para la variante del distrito de Choapan.

Se realizan en total seis pruebas las cuales consisten en analizar cada una de las realizaciones de las vocales Ę y E.

Prueba XIII. DTW aplicado a la vocal Ę en su realización cortada (Ę').

Tabla.M.14. Resultados DTW para la vocal Ę'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	22	73.333	157.324
ɐ	6	besser	Alemán	Masculino	4	13.333	174.462
ɒ	Q	lot	Ingles	Masculino	4	13.333	168.190
					30	100 %	

Prueba XIV. DTW aplicado a la vocal Ę en su realización quebrada (Ę'Ę).

Tabla.M.15. Resultados DTW para la vocal Ę'Ę.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	20	86.956	167.852
ɐ	6	besser	Alemán	Masculino	2	8.696	184.751
ɒ	Q	lot	Ingles	Masculino	1	4.348	184.862
					23	100 %	

Prueba XV. DTW aplicado a la vocal Ę en su realización sencilla (Ę).

Tabla.M.16. Resultados DTW para la vocal Ę.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	30	100	115.238
					30	100 %	

Prueba XVI. DTW aplicado a la vocal e en su realización cortada (E').

Tabla.M.17. Resultados DTW para la vocal E'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	13	61.905	168.634
ɐ	6	besser	Alemán	Masculino	4	19.048	189.513
ɒ	Q	lot	Ingles	Masculino	3	14.286	188.080
ə	@	ce	Francés	Masculino	1	4.762	189.253
					21	100 %	

Prueba XVII. DTW aplicado a la vocal e en su realización quebrada (E'E).

Tabla.M.18. Resultados DTW para la vocal E'E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɐ	6	besser	Alemán	Masculino	13	52.000	187.797
ɛ	E	gesetz	Alemán	Femenino	10	40.000	185.657
ə	@	ce	Francés	Masculino	1	4.000	193.532
ɒ	Q	lot	Ingles	Masculino	1	4.000	206.4271
					25	100 %	

Prueba XVIII. DTW aplicado a la vocal e en su realización sencilla (E).

Tabla.M.19. Resultados DTW para la vocal E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	24	100	138.557
					24	100 %	

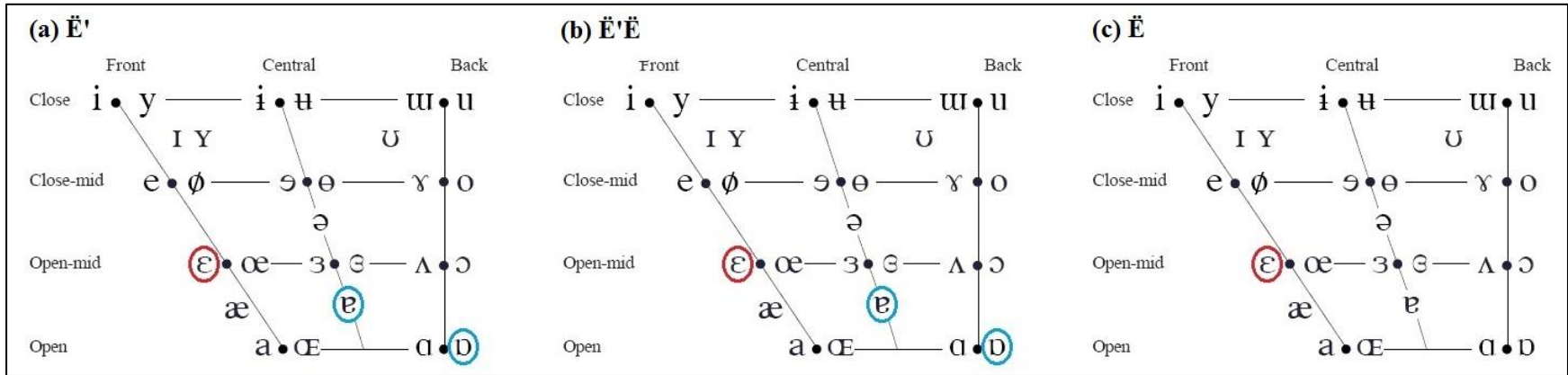


Fig.M.5. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal Ë en sus tres realizaciones para audios de prueba que pertenecen a la variante zapoteca de Choapan que se habla en el estado de Oaxaca. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

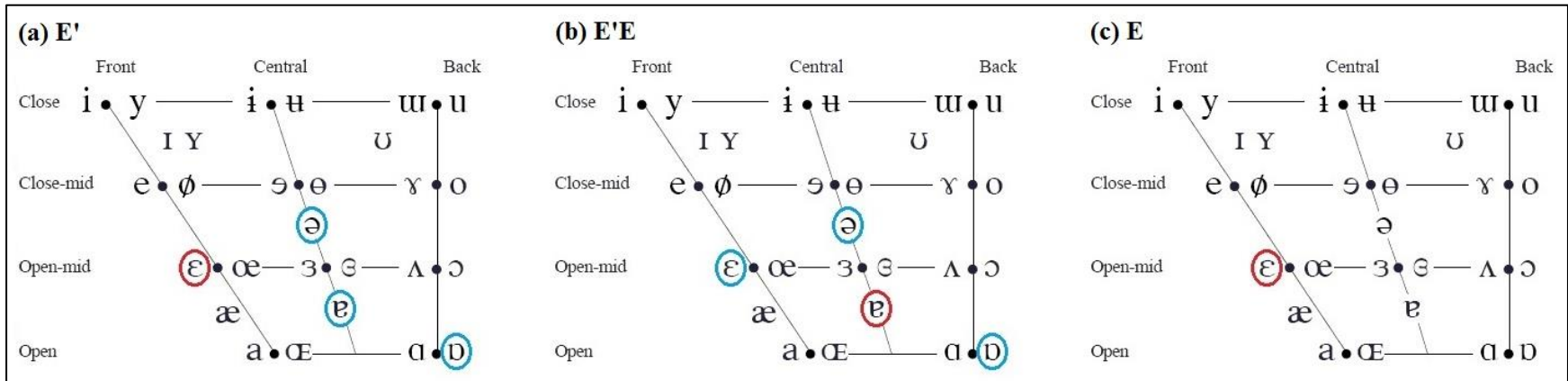


Fig.M.6. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal E en sus tres realizaciones para audios de prueba que pertenecen a la variante zapoteca de Choapan que se habla en el estado de Oaxaca. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

Resultados para la variante de Choapan del estado de Veracruz.

Se realizan en total seis pruebas las cuales consisten en analizar cada una de las realizaciones de las vocales \ddot{E} y E.

Prueba XIX. DTW aplicado a la vocal \ddot{e} en su realización cortada (\ddot{E}').

Tabla.M.20. Resultados DTW para la vocal \ddot{E}' .

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ə	@	ce	Francés	Masculino	6	42.857	203.530
Y	Y	flüsse	Aleman	Femenino	4	28.571	212.581
ε	E	gesetz	Alemán	Femenino	3	21.428	205.935
ø	8	buss	Sueco	Femenino	1	7.143	217.364
					14	100 %	

Prueba XX. DTW aplicado a la vocal \ddot{e} en su realización quebrada ($\ddot{E}'\ddot{E}$).

Tabla.M.21. Resultados DTW para la vocal $\ddot{E}'\ddot{E}$.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ε	E	gesetz	Alemán	Femenino	4	57.143	186.408
ə	@	ce	Francés	Masculino	2	28.571	209.621
ø	8	buss	Sueco	Femenino	1	14.286	201.591
					7	100 %	

Prueba XXI. DTW aplicado a la vocal \ddot{e} en su realización sencilla (\ddot{E}).

Tabla.M.22. Resultados DTW para la vocal \ddot{E} .

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ε	E	gesetz	Alemán	Femenino	4	100	203.821
					4	100 %	

Prueba XXII. DTW aplicado a la vocal e en su realización cortada (E').

Tabla.M.23. Resultados DTW para la vocal E'.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	3	30.000	224.678
ə	@	ce	Francés	Masculino	2	20.000	221.973
ɒ	Q	pot	Ingles	Femenino	2	20.000	225.637
ɒ	Q	lot	Ingles	Masculino	2	20.000	232.909
Y	Y	flüsse	Aleman	Femenino	1	10.000	243.691
					10	100 %	

Prueba XXIII. DTW aplicado a la vocal e en su realización quebrada (E'E).

Tabla.M.24. Resultados DTW para la vocal E'E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ə	@	ce	Francés	Masculino	4	66.667	209.615
ɒ	Q	lot	Ingles	Masculino	1	16.667	289.963
ɛ	E	gesetz	Alemán	Femenino	1	16.667	214.261
					6	100 %	

Prueba XXIV. DTW aplicado a la vocal e en su realización sencilla (E).

Tabla.M.25. Resultados DTW para la vocal E.

Símbolo		Palabra	Idioma	Genero	Total de Audios	Porcentaje [%]	Costo Promedio
IPA	SAMPA						
ɛ	E	gesetz	Alemán	Femenino	29	46.032	193.645
ə	@	ce	Francés	Masculino	18	28.571	196.505
ø	8	buss	Sueco	Femenino	6	9.524	275.725
Y	Y	flüsse	Aleman	Femenino	6	9.524	211.119
ɒ	Q	pot	Ingles	Femenino	4	6.349	229.174
					63	100 %	

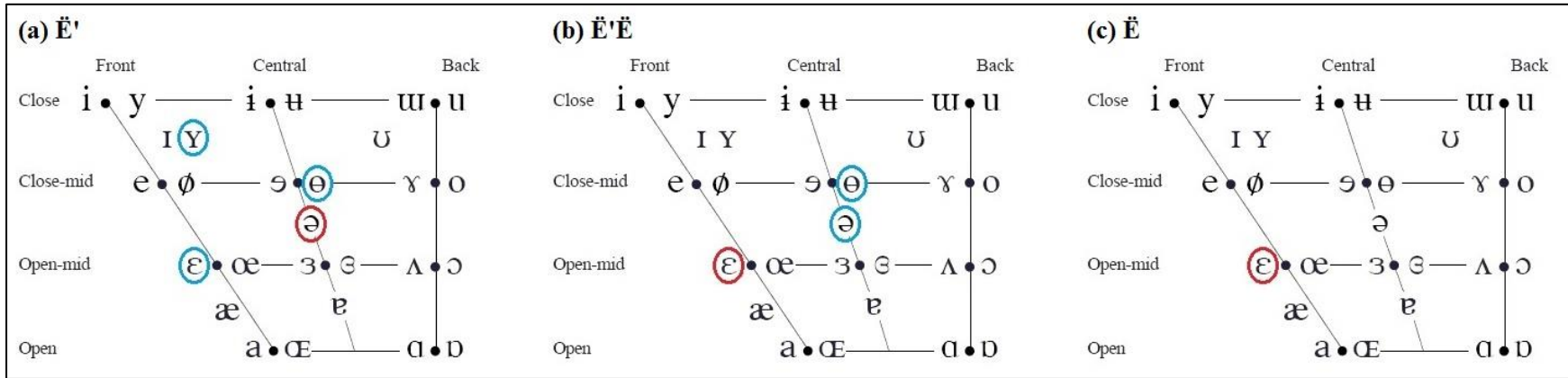


Fig.M.7. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal Ë en sus tres realizaciones para audios de prueba que pertenecen a la variante de Choapan del estado de Veracruz. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

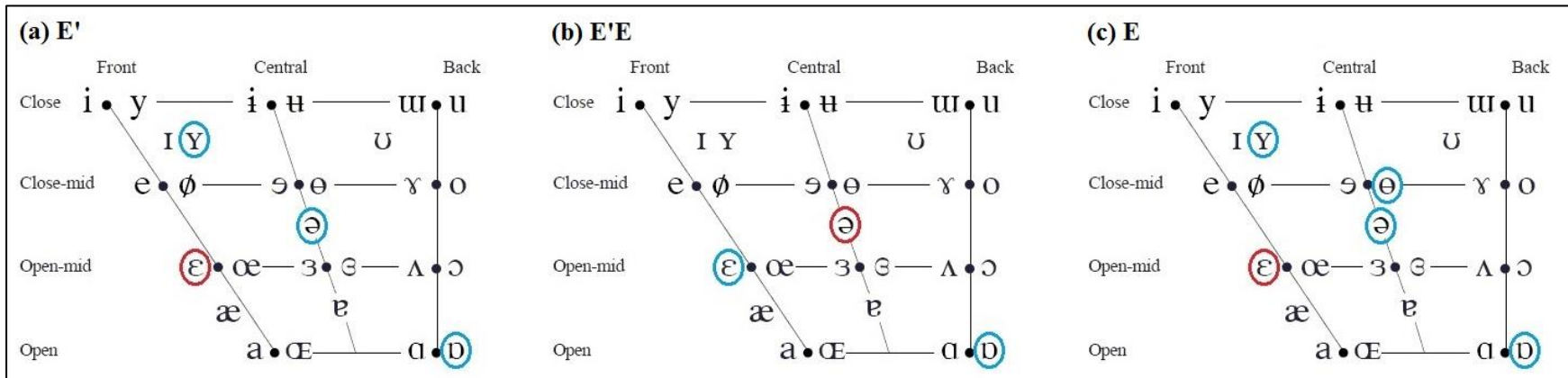


Fig.M.8. Localización de los resultados DTW dentro del Trapecio de Vocales IPA para la vocal E en sus tres realizaciones para audios de prueba que pertenecen a la variante de Choapan del estado de Veracruz. El símbolo encerrado de color rojo coincidió con la mayor cantidad de audios de prueba.

N. Artículo.

Este anexo contiene el artículo titulado “An analysis of Speech Signals of the Choapam variant Zapotec Language”, enviado al congreso SPECOM en abril de 2013.

An analysis of Speech Signals of the Choapam variant Zapotec Language

Gabriela Oliva-Juarez, Fabiola Martinez-Licona, Alma Martinez-Licona, and John
Goddard-Close

Universidad Autonoma Metropolitana, Electrical Engineering Depto., Mexico City, Mexico
{fmm1, aaml, jgc}@xanum.uam.mx, gaby7884@gmail.com

Abstract. The Zapotec language, as well as many other prehispanic languages in Mexico, is endangered for many reasons including a lack of use by the younger population who prefer to speak Spanish instead, and by the dying out of older native speakers. In this paper an analysis of the Choapam variant of Zapotec is presented; a list of words in this Zapotec was recorded, and a time and formant analysis was carried out in order to obtain basic information used to describe the language. The formant analysis focused on the vowels in the language, due complications which arise with them, and gives a first classification of them. Some of the difficulties experienced in the study, which are similar to those encountered with many other endangered languages, are detailed. Although this is a first approach to these languages analysis, it is hoped that the use of this information will contribute to further efforts aimed at helping preserve the language.

Keywords: Speech analysis, Zapotec vowels, Prehispanic languages

1 Introduction

In communication, languages present the most dynamic behavior; languages are continuously evolving, and as a consequence, some of them prevail and others disappear. The use of a particular language, and the conservation of its primary characteristics, is a critical issue towards providing a sense of identification among the communities speaking it. When the language begins to fall into disuse, there is a risk of its extinction as native speakers die off and the new generations prefer to speak in modern languages such as English or Spanish. Migration also contributes to this phenomenon. In Mexico languages like Wiyot, Mascouten, Pochuteco or Mangue have become extinct and many others are about to if there are no concerted efforts to keep them alive [1]; the Zapotec language is an example of the latter, since it is spoken by only 425,123 inhabitants [2].

Within a language there may be variants, which are variations around the standard language that identify groups of people, like families or neighbors, or even communities. Some variations are so small that are hardly noticed, and the speakers can understand one other, whilst others are extremely noticeable and can prevent an adequate communication between speakers. The Zapotec language falls into the latter category. The Zapotec language comes from the Southeast part of Mexico and it is spoken in the states of Oaxaca, Veracruz, partly in Chiapas and in a few places in Mexico City. In

Table 1. ISO 639.3 coding for some Zapotec variants

Zapotec variant	ISO639.3 Code
Asuncin Mixtepec	zpm
Choapam	zpc
Istmo	zai
Mitla	zaw
Ocotlán	zac
Rincón	zar
Zoogocho	zpq

order to identify the language and its 57 variations the ISO639.3 standard assigned a three-letter code for each one [3], and some examples are shown in Table 1.

The Zapotec language belongs to the Oto-Mangue linguistic family, which is the biggest and most diverse in Mexico. It also belongs to the linguistic group of Zapotecano that contains the macrolanguages chantino and zapoteca [1, 4]. The significant characteristic of this linguistic family is that it is tonal, meaning that it is possible to distinguish words by focusing in the frequency shifts of the vocal folds. The fact that it contains so many variants means that the Zapotec language presents an interesting and challenging problem from a speech technologies point of view. Also, the fact that the language is spoken by fewer people each year, represents an important problem for the preservation of the cultural heritage of these prehispanic communities.

In this paper an analysis of the Choapam variant of Zapotec is carried out. The speech data was collected by recording utterances of a list of words by a native speaker. The duration of the sounds involved was obtained, using the SAMPA phonetic alphabet as a reference. Due to the tonal characteristic of the language, a spectral analysis of the vowels was conducted in order to classify them. The paper is organized as follows: section 2 describes the Choapam variant of Zapotec, section 3 presents details of the development of the database as well as the words selected, section 4 reports the results of the duration of the sounds and the formant vowel analysis and classification, section 5 presents a discussion of the results and issues found during the recording process and the experimentation, and section 6 gives some conclusions and mentions further work that will be undertaken.

2 Variant of Zapotec from Choapam

The Zapotec variant of Choapam is also known as the San Juan Comaltepec Zapotec and is spoken in the municipality of Choapam, which is part of the state of Oaxaca and, according to the Summer Institute of Linguistics, also in the state of Veracruz. The data was obtained from a speaker who lives in Arenal Santa Ana, a community of 1551 inhabitants located in Veracruz [5]; although differences have been reported between the Zapotec spoken in Choapam and Arenal Santa Ana, it should be noted that due to their geographically closeness, they are highly similar. This particular variant is classified as a moribund language by the Expanded Graded Intergenerational Disruption

Scale (EGIDS) [6]. This means that the only fluent users, if any, are old and it is too late to restore the natural intergenerational transmission; it is estimated that the number of native speakers of this variant according to EGIDS is up to 10,000, as shown in Figure 1 [7]. The Choapam Zapotec alphabet consists of vowels and consonants [8], which are described in the following subsections.

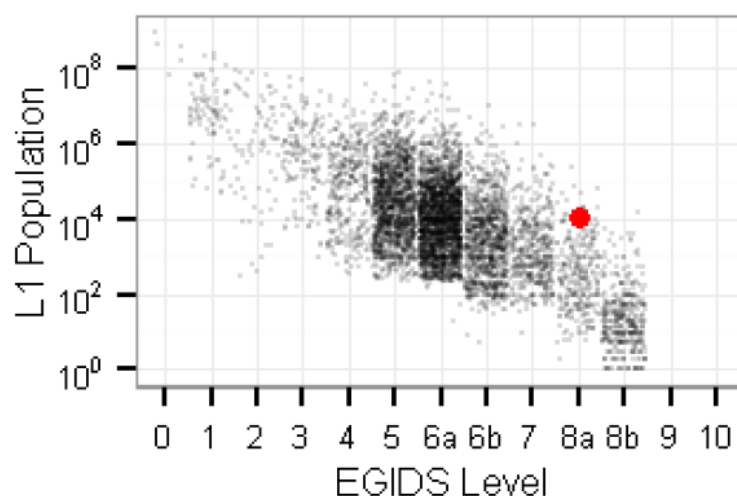


Fig.1. Location of the Choapam Zapotec within the cloud of all living languages. (Source: <http://www.ethnologue.com/cloud/zpc>)

2.1 Vowels in Choapam Zapotec

The number of vowels used in the alphabet of this variant of Zapotec is six, five of them are the same as in Spanish plus a sixth vowel, and they are spoken with three different kinds of pronunciation: simple, short and long, as shown in Table 2. The most important characteristic of the pronunciations is the vowel duration. Simple vowels have a duration that is between the short and the long vowels. These latter two pronunciations use larynx modifications to produce an abrupt closure of the vocal cords in the case of the short vowels, and to produce a brief pause between the vowel sounds in the case of long vowels.

2.2 Consonants in Choapam Zapotec

In Choapam Zapotec there are 23 consonants, most of them are similar to Spanish. The consonants are weak occlusive (*b, d, g*), strong occlusive (*p, t, c, qu, k*), weak affricative

Table 2. Vowels of Choapam Zapotec

Pronunciations	A	E	Ē	I	O	U
Simple	a	e	ĕ	i	o	u
Short	a'	e'	ĕ'	i'	o'	u'
Long	a'a	e'e	ĕ'ĕ	i'i	o'o	u'u

(*dz, dy*), strong affricative (*tz, ch*), weak fricatives (*z, ž*), strong fricatives (*s, x, j*), nasals (*m, n*), liquids (*l, r*) and semivowels (*hu, y*). The affricative sounds are not familiar sounds in Spanish, *dz* is the combination of the occlusive *d* plus the voiced *z* while *dy* sounds like a voiced *ch*; the consonant *tz* is produced by pronouncing the occlusive *t* and then a voiceless *s*. The weak fricative *ž* is not a Spanish sound, and is pronounced as a *z* but with a back position of the tongue; the rest of the phonemes are pronounced as in Spanish.

3 Database

A Mexican female from the community of Arenal Santa Ana, Veracruz was recorded uttering a list of chosen words. The list consisted of 188 words selected from the Swadesh list for Zapotec [9]. Swadesh, originally devised by the linguist Morris, contains the words that are present in almost all languages and form the basis for communication between humans. The selected words included nouns (numbers, colors, animals, parts of the body, etc.), pronouns and verbs; some examples of the words recorded and their translation to Zapotec are: heart - *lužtau'* (officially: my heart), I - *ně'di'*, dog - *bécu'*, to drink - *ra'a*, water - *nisa*, red - *na*. Each word was pronounced twice and the recording process took two sessions. The words were recorded on a desktop PC using Speech Filing System Version 4.8 [10], and a sampling frequency of 16 KHz was used. The speaker controlled the amplitude of the signals and the recording sessions were performed in a quiet room. The speaker used headphones with an integrated microphone and repeated the word to be recorded if necessary; previously she checked the pronunciation of the words by asking the elder members of her family about the non-Spanish sounds.

4 Results

The recorded words were automatically segmented using the HTK segmentation model [11]. The consonants and vowels were separated and the duration of each phoneme was extracted. Figure 2 shows the distribution of the duration, in seconds, of consonants (left) and the vowels (right). It can be seen how long vowels duration media tend to have higher values compared to the simple and short pronunciations. It also should be mentioned that the vowels short *o* (*o'*) and long *u* (*u'u*) did not appear in any word of the list, which is why there are only 16 vowels instead of 18. Also the liquid consonant *r* presented two variations: weak *r* and strong *r*, the latter was represented as *rr*.

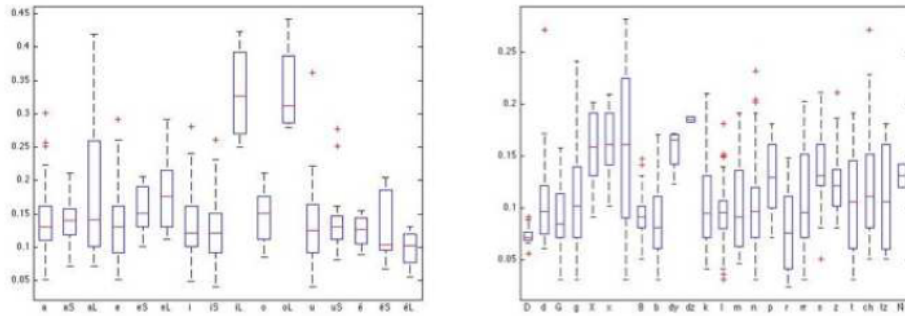


Fig. 2. Duration times of vowels (left, S:short, L:long) and consonants (right)

The vowel analysis began with the location of their first and second formant frequencies, in order to identify their properties and similarities to Spanish. Figure 3 shows the vowel location of the six vowel groups and, as expected, the formant configuration is much like those of the corresponding Spanish phonemes. In the analysis, the first five formants were extracted and two supervised learning methods were applied to the data for classification purposes. The methods used were support vector machines (SVM), using polynomial and radio basis kernels, and multilayer perceptrons (MLP), with one hidden layer and a number of hidden nodes equaling the number of classes; the results are given for a 10-fold cross validation. Table 3 shows the classification results obtained for the six vowels, and Table 4 shows the classification results of the vowels with their simple, short and long pronunciations Weka3-7-8 was used for all the experiments [12].

In the first case with six vowels, for all the methods tested, the principal conflict in the classification was between e and ë. Since the best results on the classification were obtained with MLP, the duration time was augmented to the format frequencies and the results were slightly inferior: 91.81% correct classification for MLP using formants 1 to 4 plus duration ($F_{1-4+duration}$) with 23% of corrected ë assigned, and 90.99% for MLP using formants 1 and 2 plus duration ($F_{1-2+duration}$) with 3% of corrected ë assigned. For the case of sixteen vowels, the best configuration was augmented with the duration time and it was obtained a 67.59% of correct classification on the MLP $F_{1-2+duration}$ configuration; it was not possible to improve the simple, short or long correct classification percentages since all of them appeared disperse along the time and frequency domains.

5 Discussion

The analysis of an endangered language may be difficult because of a number of different aspects that have to be taken into consideration. For example, the grammar, as well as the orthographic rules, must be known in order to form the reference frame of the study; if only a few people speak the language, some of the sounds may change or evolve as a result of the influence of foreign languages and others may disappear. The

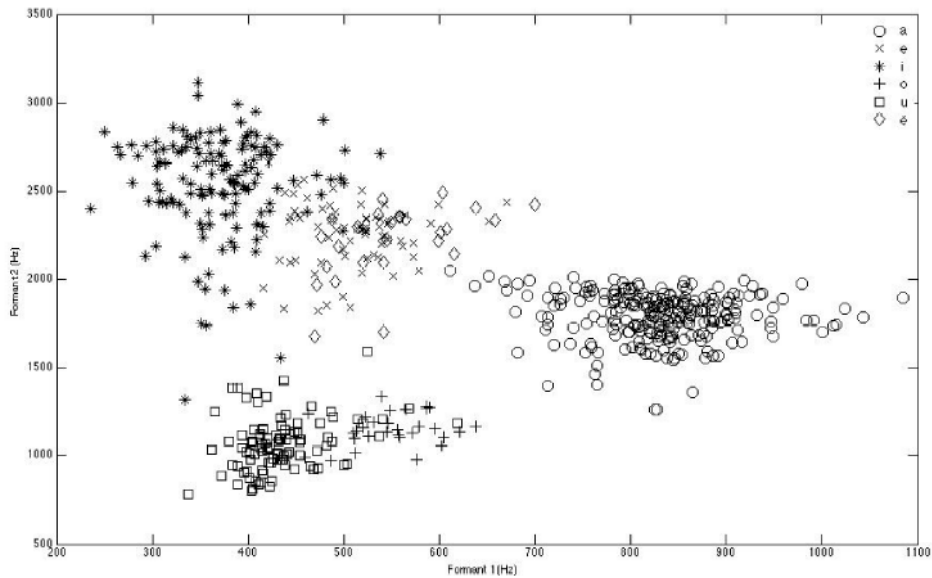


Fig. 3. Location of the first vs second formant of the Choapam Zapotec vowels

Table 3. Results of the classification of the six Choapam Zapotec vowels. SVM kernel Fi-j: Support vector Machine with kernel and formants involved, RBF Fi-j: Radio basis function with formants involved, MLP F_{i-j} : Multilayer perceptron with formants involved (6 hidden nodes, 10-fold cross validation)

Method	Correctly Classified Instances (%)	Method	Correctly Classified Instances (%)
SVM poly F_{1-5}	86.90	MLP F_{1-5}	90.34
SVM poly F_{1-4}	87.88	MLP F_{1-4}	91.98
SVM poly F_{1-3}	87.88	MLP F_{1-3}	91.32
SVM poly F_{1-2}	86.90	MLP F_{1-2}	91.98
SVM: RBF F_{1-5}	60.22		
SVM: RBF F_{1-4}	60.06		
SVM: RBF F_{1-3}	60.22		
SVM: RBF F_{1-2}	50.24		

Zapotec language, having 57 variants with varying degrees of intelligibility between them, presents a challenge to the characterization and modeling of its sounds. Some facets to consider in the development of a database, and the analysis of the speech signals, have to do with the number of examples, the type of features extracted and the realization of the Zapotec phonemes.

The words selected in the present study, belong to a generic list that was not phonetically balanced. This means that some of the phonemes have few examples, whilst others do not have any. The low classification results might improve with a larger number of

Table 4. Results of the classification of the six Choapam Zapotec vowels. SVM kernel F_{i-j} : Support vector Machine with kernel and formants involved, RBF F_{i-j} : Radio basis function with formants involved, MLP F_{i-j} : Multilayer perceptron with formants involved (6 hidden nodes, 10-fold cross validation)

Method	Correctly Classified Instances (%)	Method	Correctly Classified Instances (%)
SVM poly F_{1-5}	62.52	MLP F_{1-5}	65.13
SVM poly F_{1-4}	63.33	MLP F_{1-4}	66.12
SVM poly F_{1-3}	63.33	MLP F_{1-3}	66.77
SVM poly F_{1-2}	62.84	MLP F_{1-2}	67.26
SVM: RBF F_{1-5}	28.47		
SVM: RBF F_{1-4}	28.64		
SVM: RBF F_{1-3}	28.47		
SVM: RBF F_{1-2}	28.31		

examples for each phoneme; this leads to search the words that contain the phoneme required and record them. In a deeper study of the Choapam version of the Zapotec grammar there are few words that contain sounds like the short *o* vowel or the consonant *ž*, a careful analysis of the words, sentences and texts is advisable.

In the Choapam variant of Zapotec the problem of pronunciation is evident. The speaker had doubts about the correct spelling of certain words of the list; although a consultation with a native speaker was made, a portion of the list had to be transcribed using other variants of Zapotec like Zogocho. During the analysis of some of the voiced phonemes, in particular the case of the vowels *e* and *ě*, a mimic phenomenon could be observed: the speaker tended to pronounce the sound much like the Spanish, not allowing the differentiation of the simple, short and long pronunciations of both vowels. The classification methods employed in the paper used the formant frequencies and time durations as their inputs, however, there are other acoustic features that could also be considered, such as intensity, pitch and Mel scale cepstral coefficients. These, together with a larger dataset, could improve the classification rate obtained.

6 Conclusions

A database of Choapam variant of Zapotec was recorded, and an analysis of time and formant frequencies was applied to the vowels. The vowel sounds of this Zapotec variant are similar to the Spanish sounds and a supervised classification method of a multilayer perceptron, with one hidden layer, achieved a high correct class assignment using these features. The problems that were encountered during the speech recording dealt with the confusion on the pronunciation of some words by the speaker and members of her community, which is a consequence of the language's high risk of disappearing. Work is continuing on the recording of more speech samples in this Zapotec, the incorporation of more acoustic and spectral features such as pitch, intensity or MFCCs, and the evaluation of other classification methods. This paper presents a novel approach to the analysis of prehispanic languages using acoustic features of the speech signals.

It aims to provide useful information that help to model synthetic voices in order to preserve the language's legacy.

Acknowledgement

The first author would like to thank the Consejo Nacional de Ciencia y Tecnología (Conacyt), Mexico, for a grant given to her by them.

References

1. Catalogue of the National Indigenous Languages of Mexico, National Institute of Indigenous Languages INALI (2009) <http://www.inali.gob.mx/clin-inali/>
2. National populations statistics, National Institute of Statistics and Geography INEGI, <http://www.inegi.org.mx>
3. ISO639.3 standard, Summer Institute of Linguistics, <http://www.sil.org>
4. Ethnologue Languages of the World, <http://www.ethnologue.com>
5. Social Development Secretary, Community information modules (2010), <http://www.microrregiones.gob.mx/zap/datGenerales.aspx?entra=pdzp&ent=30&mun=130>
6. Lewis, M. Paul, & Simons, Gary. F. (2010). Assessing Endangerment: Expanding Fishman's GIDS. *Romanian Review of Linguistics*, 55(2), 103-120 (2010).
7. Zapotec, Choapan in the language cloud, <http://www.ethnologue.com/cloud/zpc>
8. Lyman-Boulden H., Gramtica Popular del Zapoteco de Comaltepec, Choapan, Oaxaca, 2a. Edicin, Summer Institute of Linguistics (2010).
9. Swadesh lists for Oto-Manguean languages, http://en.wiktionary.org/wiki/Appendix:Swadesh_lists_for_Oto-Manguean_languages
10. Speech Filing System, University College London, <http://www.phon.ucl.ac.uk/resource/sfs/>
11. Hidden Markov Model Toolkit, <http://htk.eng.cam.ac.uk/>
12. WEKA, University of Waikato <http://www.cs.waikato.ac.nz/ml/weka/index.html>

Referencias.

- [1]. Definición de lengua. Diccionario de la Real Academia Española
<http://www.rae.es/rae.html>
- [2]. Definición de lengua, dialecto e idioma, estándar ISO639.3. Summer Institute of Linguistics, <http://www.sil.org>
- [3]. Catálogo de las Lenguas Indígenas Nacionales: Variantes Lingüísticas de México con sus auto-denominaciones y referencias geoestadísticas, Instituto Nacional de Lenguas Indígenas INALI, 2009, <http://www.inali.gob.mx/>
- [4]. Familias Lingüísticas, Agrupaciones Lingüísticas, Macrolengua. Ethnologue Languages of the World, 15th. Edition, <http://www.ethnologue.com>
- [5]. Niveles de la escala EGIDS para la lengua zapoteca y sus variantes, Cantidad de lenguas en el mundo. Ethnologue Languages of the World, 16th. Edition, <http://www.ethnologue.com>
- [6]. HERRERA Z. E. y MARTIN B. P. (Eds.), “Prefacio”, En su: La tonía: dimensiones fonéticas y fonológicas”, 1a. Edición. D.F., México. Ed. El Colegio de México, 2003. pp. 9-14.
- [7]. RENDÓN M. J. J., “Notas Fonológicas del Zapoteco de Tlacochoituaaya, Anales de la Antropología”, México, Universidad Nacional Autónoma de México, Instituto de Investigaciones Históricas, v. VII, 1970, pp. 247-262.
- [8]. Datos estadísticos del censo 2010. Instituto Nacional de Estadística y Geografía, <http://www.inegi.org.mx>
- [9]. LEWIS M. P., SIMONS G. F. “Assessing Endangerment: Expanding Fishman’s GIDS”. Romanian Review of Linguistics, pp. 103-120, 2010.
- [10]. LYMAN Boulden H. “Gramática Popular del Zapoteco de Comaltepec, Choapan, Oaxaca”. 2a. Edición. D.F. Mexico. Ed. Summer Institute of Linguistics, 2010.
- [11]. RUFINER L. H. “Análisis y modelado digital de la voz. Técnicas recientes y aplicaciones”, 1ª. Edición, Argentina, Ed. Universidad Nacional del Litoral, 2009. 281pp.
- [12]. XU, C.X.; YI Xu, "F0 perturbations by consonants and their implications on tone recognition," Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on , vol.1, no., pp.I-456,I-459 vol.1, 6-10 April 2003
- [13]. “Diferencias genéticas en lenguas tonales y no tonales”, NeoFronteras, 2007. [online] <http://neofronteras.com/?p=887>
- [14]. THEERA U. N., CHANSAREEWITTAYA S., AUEPHANWIRIYAKUL S. “Phoneme and tonal accent recognition for Thai speech”, Expert Systems with Applications, Volume 38, Issue 10, pp. 13254-13259, 15 September 2011.
- [15]. SATRAVAHA N., KLINKHACHORN P., LASS N. "Tone classification of syllable-segmented Thai speech based on multilayer perceptron", System Theory, 2003. Proceedings of the 35th Southeastern Symposium on , vol., no., pp.392,396, 16-18 March 2003
- [16]. WONG Pui-Fung; SIU Man-Hung, "Integration of tone related feature for Chinese speech recognition," Signal Processing, 2002 6th International Conference on , vol.1, no., pp.476,479 vol.1, 26-30 Aug. 2002

- [17]. SCHRADER – Kniffki M., “Volumen 7 de De acá para allá: lenguas y culturas amerindias. Introducción a la lengua y cultura zapotecas”, Ed. Universidad de València, 2004
- [18]. PAFNUNCIO A. R., “Variación tonal en el zapoteco de San Pedro y San Juan Metepec, Miahuatlan, Oaxaca” in Conference on Indigenous Languages of Latin America-IV, CILLA, Austin, Texas, 2009.
- [19]. ARELLANES A. F. “Los tonos en el zapoteco de San Pablo Güilá: de la manifestación fonética a la pertinencia fonológica”, EN: HERRERA Z. E. y MARTIN B. P. (Eds.), “La tonía: dimensiones fonéticas y fonológicas”, 1a. Edición. D.F., México. Ed. El Colegio de México, 2003. pp. 37-60.
- [20]. SMITH S. T. “Tipos prosódicos de sílabas en el zapoteco de San Baltasar Chichicapan”, EN: HERRERA Z. E. y MARTIN B. P. (Eds.), “La tonía: dimensiones fonéticas y fonológicas”, 1a. Edición. D.F., México. Ed. El Colegio de México, 2003. pp. 111-139.
- [21]. SHAYAN S., OZTURK O. y SICOLI M. A., “The Thickness of Pitch: Crossmodal Metaphors in Farsi, Turkish, and Zapotec”, Senses and Society, Volume 6, issue 1, pp. 96-105, 2011.
- [22]. ADÉWÁLÉ À. L. y ÀJÀDÍ O. O., “Automatic recognition of oral vowels in tone language: Experiments with fuzzy logic and neural network models”, Applied Soft Computing, Volume 11, Issue 1, pp. 1467-1480, 2011.
- [23]. BAIRD O. B., “Un análisis acústico de las vocales del K’ichee’ ”, in Conference on Indigenous Languages of Latin America-IV, CILLA, Austin, Texas, 2009.
- [24]. GONZÁLEZ R. M. C. y ESTRADA H., “Aproximación fonético - espectrográfica de algunos aspectos de la fonología sáliba”, Lingüística y Literatura, No. 54, 2008. [online] <http://aprendeenlinea.udea.edu.co/revistas/index.php/lyl/>
- [25]. STOBART H. y CROSS I., The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosi, Bolivia, British Journal of Ethnomusicology, Vol.9, No.2, pp. 63-92, 2000.
- [26]. NOLAZCO F.J.A., SALGADO G.L.R. y PEÑA D. M., “Speaker dependent ASRs for huastec and western-huastec nahuatl languages”, Lecture Notes in Computer Science, Vol. 3523, pp. 595-602, 2005.
- [27]. STOBART H. y CROSS I., The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosi, Bolivia, British Journal of Ethnomusicology, Vol.9, No.2, pp. 63-92, 2000.
- [28]. Audio, Libro. Scripture Earth, Recursos Bíblicos. http://www.scriptureearth.org/00e-Escrituras_Indice.php?sortby=country&name=all
- [29]. MAZZONI D. Y DANNENBERG R., Audacity®, Universidad de Carnegie Mellon, <http://audacity.sourceforge.net/?lang=es>
- [30]. ODDEN David, “Introducing Phonology”, New York. Ed. Cambridge University Press, pp. 360, 2005
- [31]. OBEDIENTE E. “Fonética y Fonología”, 3a. Edición. Venezuela, Ed. Consejo de Publicaciones de la Universidad d los Andes, 391pp, 2007.
- [32]. SAMPA - computer readable phonetic alphabet. <http://www.phon.ucl.ac.uk/home/sampa/>

- [33]. Li, B.N.L.; Liu, J.N.-K., "A comparative study of speech segmentation and feature extraction on the recognition of different dialects," Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on , vol.1, no., pp.538,542 vol.1, 1999
- [34]. JITAPUNKUL S., LUKSANEYANAWIN S., AHKUPUTRA V., MANEENOI E., KASURIYA S. y AMORNKUL P., "Recent advances of Thai speech recognition in Thailand," Circuits and Systems, 1998. IEEE APCCAS 1998. The 1998 IEEE Asia-Pacific Conference on, pp.173,176, 24-27, 1998.
- [35]. Satravaha, N.; Klinkhachorn, P.; Lass, N., "Tone classification of syllable-segmented Thai speech based on multilayer perceptron," System Theory, 2003. Proceedings of the 35th Southeastern Symposium on , vol., no., pp.392,396, 16-18 March 2003
- [36]. Hidden Markov Model Toolkit (HTK). <http://htk.eng.cam.ac.uk/>
- [37]. YOUNG S., EVERMANN G., GALES M. y otros: "The HTK Book (for HTK Version 3.4)", Ed. Cambridge University Engineering Department, 2006. 368 pp.
- [38]. BOERSMA P. y WEENINK D., Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>
- [39]. GORMAN K., "Automatic speech segmentation with HTK". [online] <http://www.ling.upenn.edu/~kgorman/papers/segmentation/.speechseg.html>
- [40]. MARLETT A. S. "Basic Vocabulary", SIL, 2010, 66p.
- [41]. LONG C. R. y CRUZ M. S., "Diccionario Zapoteco de San Bartolomé Zoogocho. Oaxaca", 2ª. Edición, Mexico, SIL, 2000. 549 pp.
- [42]. RABINER L. y JUANG B. "Fundamentals of speech Recognition", E. U., Ed. Prentice Hall, 1993, 507 pp.
- [43]. MÜLLER M. "Information Retrieval for Music and Motion", Ed. Springer. 2007. 318pp.
- [44]. The MathWorks, Inc. Matlab R2012b.
- [45]. Dynamic Time Warp (DTW) in Matlab. <http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/>
- [46]. WEKA, University of Waikato. <http://www.cs.waikato.ac.nz/ml/weka/index.html>
- [47]. Weenink D. "Speech Signal Processing with Praat", 2013, 328pp. [online] <http://www.fon.hum.uva.nl/david/sspbbook/sspbbook.pdf>
- [48]. LADEFOGED P. "Phonetic Data Analysis. An Introduction to Fieldwork and Instrumental Techniques", Ed. Blackwell Publishing. 193 pp.
- [49]. BOERSMA P. "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound", Institute of Phonetic Sciences, University of Amsterdam , pp. 97-110, 1993. [online] http://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf
- [50]. ALI A., BHATTI S., MIAN M.S., "Formants Based Analysis for Speech Recognition," Engineering of Intelligent Systems, IEEE International Conference on , vol., no., pp.1,3, 2006.
- [51]. Manual, ©ppgb, 2010. [online] <http://>

http://www.fon.hum.uva.nl/praat/manual/Sound__To_Formant__burg____.html

- [52]. RUSSELL S. y NORVIG P. “Inteligencia Artificial. Un Enfoque Moderno”, 2ª. Edición, España, Ed. PEARSON Prentice Hall, 2004. 1212 pp.
- [53]. THE INTERNATIONAL PHONETIC ASSOCIATION, “Handbook of the International Phonetic Association”, E.U., Ed. Cambridge University Press, 2007. 204 pp.
- [54]. THE INTERNATIONAL PHONETIC ASSOCIATION, Alfabeto Fonético Internacional. <http://www.langsci.ucl.ac.uk/ipa/>
- [55]. WELLS J.C., “Computer- coding the IPA: a proposed extension of SAMPA”, University College London. [online] <http://www.phon.ucl.ac.uk/home/sampa/ipasam-x.pdf>
- [56]. Forvo, All the words in the world. Pronounced. <http://es.forvo.com/languages/>
- [57]. MARIANI J. “Language and Speech Processing”, 1ª. Edición, Ed. ISTE Publishing Company, 2008. 416pp.
- [58]. INSTITUTO NACIONAL DE ESTADISTICA Y GEOGRAFIA. “Perspectiva estadística. Veracruz de Ignacio de la Llave”, INEGI, 2012. [online] http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/integracion/estd_perspect/ver/Pers-ver.pdf
- [59]. Academia Veracruzana de Lenguas Indígenas (AVELI), “Catálogo de Lenguas Indígenas y sus Variantes Lingüísticas del Estado de Veracruz”, 2010. [online] www.aveli.gob.mx/
- [60]. Secretaría de Finanzas y Planeación del Estado de Veracruz (SEFIPLAN), “Sistema de Información Municipal. Playa Vicente. Cuadernillo municipal”. SEFIPLAN, 2013. [online] <http://www.veracruz.gob.mx/finanzas/files/2013/04/Playa-Vicente.pdf>
- [61]. Arenal Santa Ana, Playa Vicente, Veracruz-Llave. 2013., Oliva Juárez Gabriela, “Google Maps”, (https://www.google.com.mx/mapmaker?ll=17.857212,-95.908585&spn=0.729375,1.71524&t=h&z=10&q=Arenal+Santa+Ana,+Playa+Vicente,+Veracruz-Llave&utm_source=mapseditbutton_normal&gw=30&lyt=large_map_v3&hyaw=42.99818819577434).
- [62]. Swadesh lists for Oto-Manguean languages. http://en.wiktionary.org/wiki/Appendix:Swadesh_lists_for_Oto-Manguean_languages



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No. 00040

Matrícula: 2113802361

ANALISIS DEL ZAPOTECO
MEDIANTE TECNICAS DE
RECONOCIMIENTO DE PATRONES

En México, D.F., se presentaron a las 13:00 horas del día 31 del mes de enero del año 2014 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. IRINEO LORENZO LOPEZ CRUZ
DR. JOHN CHARLES HENRY GODDARD CLOSE
M. EN C. ALMA EDITH MARTINEZ LICONA



GABRIELA OLIVA JUAREZ
ALUMNA

Bajo la Presidencia del primero y con carácter de Secretaria la última, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRA EN CIENCIAS (CIENCIAS Y TECNOLOGIAS DE LA INFORMACION)

DE: GABRIELA OLIVA JUAREZ

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

Aprobar

Acto continuo, el presidente del jurado comunicó a la interesada el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

REVISO

LIC. JULIO CESAR DE LARA ISASSI
DIRECTOR DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JOSE ANTONIO DE LOS REYES
HEREDIA

PRESIDENTE

DR. IRINEO LORENZO LOPEZ CRUZ

VOCAL

DR. JOHN CHARLES HENRY GODDARD
CLOSE

SECRETARIA

M. EN C. ALMA EDITH MARTINEZ LICONA