Capítulo 5.

Presentación de los resultados

Es necesario efectuar una evaluación final del comportamiento del reconocedor y del método de verificación de pronunciación para medir su desempeño.

5.1 Experimento 1

En el primer experimento, se trabajó con un *script* que reconocía lo que la persona dijo utilizando un vocabulario y archivo de especificación de partes basados en fonemas (ver Apéndice C). Se aplicó utilizando un reconocedor de propósito general dependiente de contexto a nivel fonético desarrollado por el Grupo Tlatoa.

La salida generada fue del tipo:

Tabla 5.1 Ejemplos de salida del primer experimento

Palabra	Salida
Abeja	.pau a D j dZ e i s tS f e s kc dc k
Beber	pc .pau bc dZ e V e rr r e s a .pau
Cebolla	.pau s e D o j dZc dZ j es b .pau
Chocolate	.pau tSc tS o pc kc u j l a pc bc f dc dZ i .pau
Delfín	.pau dc pc dZ bc dZ e i s f i u e n m f tc p .pau
Fruta	.pau tSc tS u rr o pc tc r dZ a a .pau
Gato	k .pau dZc dc r bc rr dZc gc dZc dZ r dZ a pc tc pc tc f dZ
Hola	o l a s x s j n N pc dZc pc k
Limpiar	.pau bc dZc bc dZ i N n tc pc r dZ i a r a .pau
Maíz	.pau bc m rr m a e i s bc m

Como se puede apreciar, este experimento arrojó datos que no satisfacían nuestros requerimientos, es decir, se reconocía mucho más de lo que se había dicho o no se reconocían correctamente los fonemas y por tanto no podía utilizarse para los fines requeridos.

5.2 Experimento 2

En el segundo experimento se trabajó con un *script* que reconocía lo que la persona dijo, en este caso se utilizó un vocabulario y archivo de partes basados en palabras y para cada palabra su pronunciación, la gramática estaba basada en las palabras contenidas en el vocabulario y las pronunciaciones de las mismas. El reconocedor que se utilizó fue de propósito general, dependiente del contexto a nivel fonético desarrollado por el Grupo Tlatoa. La salida obtenida fue del tipo:

Tabla 5.2 Ejemplos de salida del segundo experimento

Palabra	Salida
Abeja	.pau a bc b e x a .pau
Beber	.pau bc b e bc b e r .pau
Cebolla	.pau s e bc b o ll a .pau
Chocolate	.pau tSc tS o kc kc o l a tc t e .pau
Delfín	.pau dc d e l f i n .pau
Fruta	.pau f r u tc t a .pau
Gato	.pau gc g a tc t o .pau
Hola	.pau o l a .pau
Limpiar	.pau l i m pc p i a r .pau
Maíz	.pau m a i s .pau

La salida era buena, aunque no solucionaba el problema de pronunciación pues lo que hacía era mapear una secuencia de fonemas a una palabra y no mostraba realmente los errores. Por ejemplo:

```
PALABRA A VERIFICAR: Abeja
EL ESTUDIANTE DIJO: /ei/ /b/ /e/ /j/ /a/
SALIDA: pau a bc b e x a .pau
```

```
PALABRA A VERIFICAR: Abeja
EL ESTUDIANTE DIJO: /e/ /b/ /ei/ /j/ /a/
SALIDA: pau a bc b e x a .pau
```

```
PALABRA A VERIFICAR: Abeja
EL ESTUDIANTE DIJO: /ae/ /b/ /e/ /j/
SALIDA: pau a bc b e x a .pau
```

Este experimento no satisfacían nuestros requerimientos. La siguiente opción era definir todas las posibles pronunciaciones para una palabra. Este proceso trajo consigo varios problemas.

Primero: la definición para cada palabra, ¿cuáles eran todas las posibles pronunciaciones que el estudiante podía decir para la palabra?. Las

posibilidades eran muchas. Y, entre más fonemas tuviera una palabra más posibles pronunciaciones había.

Segundo: la búsqueda de la pronunciación correcta dentro de las posibles pronunciaciones. Debido a que el espacio de búsqueda era muy grande el tiempo de respuesta aumento. Y, a pesar de existir diferentes pronunciaciones para la palabra, la salida del reconocedor no era siempre correcta.

Tercero: el tamaño del vocabulario. Entre más palabras para verificar se tuvieran más definiciones se tenían que hacer en el archivo .vocab.

Esta modificación a el experimento satisfacía algunos los requerimientos, pero no era la forma más adecuada de hacerlo.

5.3 Experimento 3

El siguiente experimento se hizo tomando la idea anterior, definición de posibles pronunciaciones, pero en vez de basarnos en palabras nos basamos en fonemas.

No se definieron todas las posibles opciones para cada fonema, si no que solo se definieron las pronunciaciones más frecuentes. Para hacer tal definición se contó con el apoyo de profesores del Departamento de Lenguas de la Universidad de las Américas que imparten cursos de Español para extranjeros.

Para el proceso de verificación, se trabajo con un *script* que alinea lo que la persona dijo a un texto, genera una gramática y luego lleva a cabo el proceso de reconocimiento, utilizando un vocabulario basado en fonemas. Se aplicó después de entrenar un reconocedor de propósito general dependiente del contexto con 400 locutores a nivel fonético.

El script se desarrolló en tcl, y se utiliza de la siguiente forma:

```
tcl verifica.tcl "u n o" uno.wav salida.out
```

donde tcl es el interprete, verifica.tcl es el nombre del script que se va a ejecutar, "u n o" es el texto en el que se basa para alinear, uno.wav es el archivo de sonido .wav (pronunciación a evaluar) y salida.out es el archivo de salida.

Para ejecutar el script es necesario tener una red neuronal entrenada, un archivo .vocab, un archivo .olddesc el cual es generado por el proceso de entrenamiento, el texto con que quieres forzar tu archivo .wav, el archivo .wav y un nombre para tu archivo de salida.

En la ejecución del script de verificación (ver Apéndice D), primero se crea la gramática para la palabra o frase a evaluar. Se buscan las pronunciaciones para cada fonema leyendo el archivo .vocab. y se construye la gramática de acuerdo a los fonemas de la palabra o frase a evaluar. Luego se lleva a cabo la búsqueda para saber cual fue el fonema que realmente pronunció, para tal proceso primero se lee el

archivo .wav, se crea el vector de características en base a este archivo, crea la matriz de probabilidades, hace la actualización de la puntuación de los estados y obtiene los resultados. Y finalmente guarda el resultado en un archivo de salida.

La salida generada por éste script fue del tipo:

Tabla 5.3 Ejemplos de salida para el experimento 3

Palabra	Salida
Abeja	.pau a bc b ei x a .pau
Beber	.pau bc b i: bc b e r .pau
Cebolla	.pau s i: bc b ow ll a .pau
Chocolate	.pau tSc tS o kc kc ow l a tc t e .pau
Delfín	.pau dc d e l f ie n .pau
Fruta	.pau f r ow tc t a .pau
Gato	.pau gc g au tc t ow .pau
Hola	.pau ow l a .pau
Limpiar	.pau l i m pc p i a 9r .pau
Maíz	.pau m a i s .pau

Este experimento satisface nuestros requerimientos. Reconoce lo que la persona dijo basándose en una gramática restringida.

Existe una técnica, que se aplicó a este caso, que sirve para verificar el desempeño del reconocedor. La técnica consiste en comparar los resultados dados por el reconocedor, con los resultados humanos, es decir, varias personas analizan cada frase e indican donde hubo error de pronunciación [Bernstein, 97; Cucchiarini, et al., 99; Franco, et al., 97; Witt & Young, 97].

Para aplicar la técnica citada, se desarrolló una aplicación en el RAD mediante la cual las personas evaluadoras escucharon las frases grabadas por los estudiantes norteamericanos y llenaron una hoja de evaluación para cada uno (ver Apéndice C). En total fueron 15 personas que evaluaron el corpus de pruebas. La aplicación puede verse en la siguiente figura.

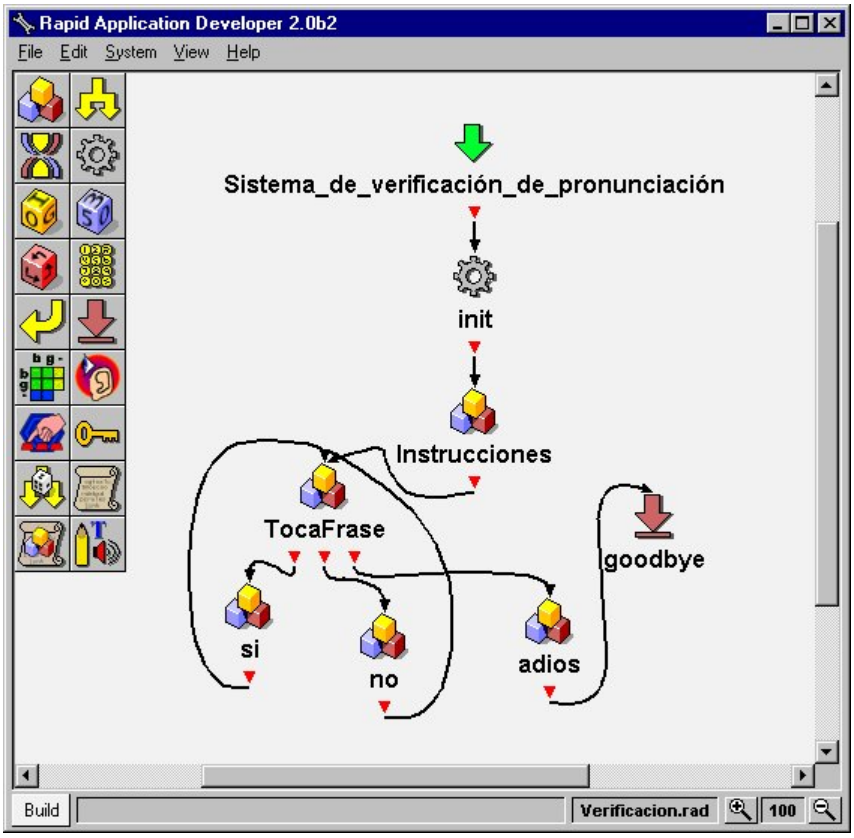

Fig 5.1 Aplicación para la evaluación humana

El script de verificación se probó en el corpus de pruebas, el cuál fue también evaluado por el juicio humano. En la tabla 5.4 podemos apreciar los resultados arrojados por las pruebas de verificación (juicio humano y reconocedor) para los errores marcados sobre los fonemas.

Tabla 5.4 Comparación entre el juicio humano y la salida del método de verificación

Frase	El juicio humano	El reconocedor
Botella Rápido Traer Refresco	El 12.8% marcó errores con las 'r', un 50.66% errores en la 'll', un 3.5% errores en la 'o', el 0.88% errores en 'a' y 0.88% errores en 'e'.	Marcó error en la 'll' en un 80% de las veces. Marcó error con la 'r' en un 10% de las veces. Marcó errores con vocales que no fueron consideradas como tal.
Comer Cenar Dulce Tomar	El 2.6% marcó errores en la 'c', el 14% errores en la 'r', el 10.66% errores en la 'l' y un 1.3% error en la 'u' de dulce.	No marcó ningún error con la 'r' y con la 'c', marcó error el 100% de las veces al igual que con la 'u'.
Café Te Azúcar Pan	Un 4.33% marcó errores en la 'a', el 10.66% errores en la 'z' y un 8% errores con la 'r'.	No marcó ningún error con la 'z' y con la 'r'. Marcó error en la 'a' y marcó error con la 'u'.

Crema Chile Frío Caliente	El 4% marcó errores con la 'e', un 2% en la 'l', un 0.66% en la 'r', un 0.13% en la 'o' y un 0.13% en la 'ch'.	No marcó ningún error.
Bien Pastel Flan Pay Chocolate	El 16% marcó errores con el fonema 'ay', el 2% errores con la 'o' y el 5.3% errores con la 'ch'.	Marcó el error en 'ay' el 83% de las veces, no marcó los errores de la 'ch' y la 'o' y marcó error en la 'e'.
Fresa No Si Salud	El 5.33% marcó errores en 'r', 14.6% errores en la 'o' y 4% en el fonema 'e'.	No marcó errores en la 'r', 'o' y 'e'. Marcó errores en la 'u' y en la 'a' de sal.
Vegetariano Cuenta Cerveza Bebida	El 0.53% marcó errores en la 'a', 1.77% en la 'e', 1.33% errores en la 'r' y 2.6% errores en la 'z' .	Marcó error en la 'e' en un 100% de las veces. No marcó error en la 'a', 'r' y 'z' y marcó error en la 'v' de vegetariano.
Jugo Agua Vino Sal	Un 0.44% marcó error en la 'a', un 0.66% en la 'u' y un 0.66% en la 'l'.	No marcó ningún error.
Pimienta Carne Res Puerco	Un 1.33% marcó errores con la 'r'.	No marcó ningún error.
Vaca Pescado Aves Verduras	3.33% marcó errores con la 'r', un 0.44% con la 's' y un 0.88% con la 'v'.	Marcó errores con la 'a', y con la 'u'
Fruta Papa Ensalada Postre	El 0.08% marcó errores con la 'r', el 4% con la 'o' y el 0.66% con la 'a'.	Marcó al 100% los errores con la 'a' y marcó errores con la 'u'.
Helado Tortilla Sabor Cebolla	El 21.22% marcó errores con la 'll' y el 2.6% errores con la 'o'.	Marcó el error de la 'll' en un 40% de las veces y marcó errores con la 'a', la 'e' y la 'i'.

Como mencionamos en capítulo anterior el corpus de pruebas consta de las grabaciones hechas por cinco locutores, en la tabla 5.5 podemos ver las estadísticas determinadas para el locutor número 2, dichas estadísticas se realizarón para cada locutor. Por cada frase se determinó el número total de fonemas. Con respecto a la evaluación humana tenemos el total de errores marcados, el porcentaje de error y el porcentaje de evaluadores que marcó el error. Con respecto a la evaluación automática tenemos el total de errores, el porcentaje que éstos representan, el porcentaje de concordancia con respecto a la evaluación humana, el total de errores que el sistema marcó y el humano no marcó, el porcentaje que representa, el total de errores que el sistema no marcó y que el humano marcó y el porcentaje que

representa.

Tabla 5.5 Estadísticas por locutor

		Eval. humana			Eval. Automatica con respecto a la humnana						
Frases	# fonemas	errores	% de error	% evaluadores	error	% error	Concuerdan	Err. Marcado	Err. Marcado	Err. Omitido	Er Or
Frase 1	25	5	20.00%	100.00%	3	12.00%	92.00%	0	0.00%	2	8.(
Frase 2	20	1	5.00%	13.33%	0	0.00%	95.00%	0	0.00%	1	5.(
Frase 3	15	1	6.67%	13.33%	1	6.67%	86.67%	1	6.67%	1	6.(
Frase 4	21	0	0.00%	0.00%	0	0.00%	100.00%	0	0.00%	0	0.
Frase 5	24	1	4.17%	20.00%	2	8.33%	95.83%	1	4.17%	0	0.
Frase 6	14	1	7.14%	20.00%	2	14.29%	78.57%	2	14.29%	1	7.1
Frase 7	30	1	3.33%	20.00%	3	10.00%	93.33%	2	6.67%	0	0.
Frase 8	15	1	6.67%	20.00%	0	0.00%	93.33%	0	0.00%	1	6.(
Frase 9	22	1	4.55%	46.67%	0	0.00%	94.45%	0	0.00%	1	4.!
Frase 10	23	0	0.00%	0.00%	1	4.35%	95.65%	1	4.35%	0	0.
Frase 11	23	1	0.00%	0.00%	1	4.35%	95.65%	1	4.35%	0	0.
Frase 12	24	1	4.17%	20.00%	2	8.33%	95.83%	1	4.17%	0	0.
Total	256	13	5.08%		15	5.86%	93.75%	9	3.52%	7	2.7

En resumen tenemos la siguiente tabla donde podemos apreciar para cada estudiante el número de fonémas pronunciados. En cuanto a la evaluación humana tenemos el total de errores marcados y el porcentaje que representa. Con respecto a la evaluación automática se da el número de errores y el porcentaje que representa, el porcentaje de fonemas que concuerdan, el total de fonemas erroneamente marcados y el porcentaje que representan, el total de fonemas erroneamente omitidos y el porcentaje que representa.

Tabla 5.6 Resumen de las estadísticas

		Eval. humana		Eval. . Automática con respecto a la humana						
	#fonemas	errores	% de error	errores	% de error	Concuerdan	Err. Marcado	Err. Marcado	Err. Omitido	Err. Omitido
Estudiante 1	256	39	15.23%	15	5.86%	83.59%	9	3.52%	33	12.89%

Estudiante 2	256	13	5.08%	15	5.86%	93.75%	9	3.52%	7	2.73%
Estudiante 3	256	15	5.86%	9	3.52%	92.97%	6	2.34%	12	4.69%
Estudiante 4	256	20	7.81%	15	5.86%	87.89%	13	5.08%	18	7.03%
Estudiante 5	256	21	8.20%	9	3.52%	89.84%	7	2.73%	19	7.42%
Total	1280	108	8.44%	63	4.92%		44	3.44%	89	6.95%

Conclusión

Como podemos observar el desempeño del reconocedor es bueno, pero tiene muchos problemas en reconocer algunos fonemas con problemas como la 'r', la 'o' y la 'z'.

Cabe hacer mención que muchas veces el reconocedor marcó errores que el juicio humano no vio.

Es de importancia notar que cuando hablamos no pronunciamos perfectamente todos los fonemas. Es decir, al pronunciar una palabra tenemos errores que no notamos, por ejemplo la palabra 'puerco' es muy común que sea pronunciada como 'poerco'.

Aguas García, N. 1999. Verificación de Pronunciación Basada en Tecnología de Reconocimiento de Voz para un Ambiente de Aprendizaje. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Diciembre.