ELSA MORA*
LOURDES PIETROSEMOLI*
CHRISTIAN CAVÉ**
ENRIQUE OBEDIENTE*
ERWIN LA CRUZ***
Departamento de Lingüística, Universidad de los Andes, Mérida, Venezuela*
Laboratoire Parole et Langage, Université de Provence, France**
Escuela de Idiomas Modernos, Universidad de los Andes, Mérida, Venezuela***
elsamora@icnet.com.ve, lourdes@ula.ve, christian.cave@lpl.univ-aix.fr, enriobe@cantv.net
erwinlacruz@yahoo.co.uk

Un corpus de pares mínimos para el español de Venezuela[1]

1. Introducción

Las pruebas de pares mínimos (PPM) sobre todo en sus formas de pruebas de rima, constituyen una herramienta excelente para la evaluación de la inteligibilidad del habla a pesar del desarrollo de otros procedimientos más recientes [1, 2]. De hecho, las PPM no solamente proporcionan una medida global de la tasa de inteligibilidad sino también un diagnóstico sobre el tipo de error realizado por un sistema dado. La PPM aquí presentada ha sido desarrollada, al igual que otros corpus que serán presentados posteriormente, de manera de poder evaluar el rendimiento de un sistema de síntesis de voz para el español venezolano que se está desarrollando en los actuales momentos en nuestro laboratorio [3]. El principio de una prueba de pares mínimos consiste en oponer dos unidades léxicas que se diferencian la una de la otra por un solo rasgo y en evaluar la inteligibilidad de estas dos unidades bajo condiciones determinadas. Nos interesa particularmente en este trabajo el nivel de los rasgos distintivos de los fonemas [4] lo que nos da la posibilidad de analizar las confusiones fonéticas a un nivel infrafonémico. Es de hacer notar que esta prueba, desarrollada inicialmente para evaluar un sistema de síntesis de voz, puede también convertirse en una herramienta importante de valoración de la audición humana en ambientes clínicos, como el test presentado por Cárdenas y Marrero [5]

2. El Corpus

Tomando como base la descripción fonológica de las consonantes del español venezolano presentada por Obediente [6], se escogieron siete rasgos que permiten describir el sistema consonántico del español venezolano. Se ha escogido, como suele hacerse para las pruebas de rima, examinar los rasgos consonánticos en posición inicial y en todos los entornos vocálicos, pero teniendo en cuenta las restricciones léxicas de la lengua. Esto impide que se recurra a logotomos o a palabras inexistentes para examinar secuencias consonante-vocales que no sean autorizados o que no existan en el léxico. En consecuencia, el rasgo tenso/laxo que no se presenta en posición inicial de palabra se ha examinado en posición intervocálica como en cerro/cero.

El corpus consta de 74 pares de palabras que constituyen un tipo especial de pares mínimos: los fonemas que definen los pares en cuestión difieren en un solo rasgo distintivo. Las palabras son en su mayoría bisílabas por ser la estructura bisilábica más frecuente en español que la monosilábica. En este sentido, la prueba difiere de otras similares hechas para el inglés o el francés. En un esfuerzo por controlar al máximo las respuestas constituidas por la presentación de los pares mínimos, el presente corpus se elaboró presentando los contrastes sólo al comienzo de palabra y evaluando estos contrastes en los cinco entornos vocálicos fonológicos del español. Esta condición no ha podido cumplirse para todos los rasgos examinados debido a restricciones impuestas por la misma lengua como en el caso del rasgo tenso que diferencia las vibrantes múltiple y simple. En este caso, se ha recurrido a la posición intervocálica.

2.1. *Principio de constitución*

Hemos seleccionado los siete rasgos que se ejemplifican a continuación:

Sonoro/sordo	b/p
Grave/agudo	p/t
Compacto/difuso	k/p
Interrupto/No interrupto	rr/l
Tenso/relajado	rr/r
Vocálico/No vocálico	l/d
Nasal/No-nasal	m/b

Teniendo en cuenta que es preferible evaluar un rasgo con varias consonantes diferentes y las restricciones léxicas mencionadas, presentamos a continuación la tabla de pares mínimos.

[1] Agradecimientos: Este trabajo ha sido realizado con el apoyo de Ecos-Nord (Acción V99H01) y de Fonacit (Código 99000240), Venezuela.

Tabla 1. Lista de pares mínimos propuesta para el español venezolano

Sonoro	**a**	**e**	**i**	**o**	**u**
b p	bata pata	besa pesa	bies pies	bolo polo	bus pus
d t	dardo tardo	deme teme	día tía	dos tos	duna tuna
g k	gato cato	----------	guiso quiso	goza cosa	gurí curí
Grave	**a**	**e**	**i**	**o**	**u**
f s	falta salta	feria seria	fino sino	fonda sonda	fuma suma
p t	pala tala	pela tela	pino tino	poro toro	pus tus
b d	bata data	bebe debe	bizco disco	bote dote	buche duche
m n	mata nata	meto neto	mido nido	mota nota	mudo nudo
m ñ	mato ñato	meco ñeco	minga ñinga	moño ñoño	mu ñu
Compacto	**a**	**e**	**i**	**o**	**u**
k p	carta parta	queso peso	quita pita	coro poro	cura pura
g b	gala bala	guerrear errear	guía vía	gota bota	gusto busto
Interrupto	**a**	**e**	**i**	**o**	**u**
rr l	rama lama	reza lesa	risa lisa	roto loto	rudo ludo
Tenso	**a**	**e**	**i**	**o**	**u**
rr r	parra para	cerro cero	cirro Ciro	morro moro	curro curo
Vocálico	**a**	**e**	**i**	**o**	**u**
l d	lato dato	lejos dejos	lime dime	lomo domo	luna duna
Nasal	**a**	**e**	**i**	**o**	**u**
m b	mata bata	mesa besa	millar billar	mote bote	mula bula
n d	nado dado	neceo deseo	Nilo dilo	nota dota	nudo dudo

3. Grabación y evaluación

3.1 *Grabación*

El corpus fue grabado en sala insonorizada por una hablante nativa. La grabación se realizó directamente en el disco duro de un computador y simultáneamente en un DAT (Casete de audio digital).

3.2 Validación

El corpus se presentó, palabra por palabra, a un grupo pequeño de oyentes nativos a un nivel medio de 65 dBA. Los sujetos presentaron sus respuestas en una hoja de respuestas al escuchar la palabra oída. En estas condiciones, la tasa de inteligibilidad alcanzó un 100% y ningún ítem fue afectado por errores sistemáticos.

4. Conclusión

Teniendo en cuenta las condiciones de grabación y los resultados de esta primera evaluación, se puede considerar que la calidad acústica y fonética del corpus es satisfactoria y que el mismo puede utilizarse para la recolección de datos de referencia que sirvan para evaluar el rendimiento de los sistemas de síntesis de voz o de otros dispositivos tales como las prótesis auditivas o los implantes cocleares. Sin embargo, la versión que presentamos aquí es provisional y podrá ser modificada y optimizada en función de los resultados de evaluaciones más avanzadas y con un número más representativo de hablantes nativos, lo que se está llevando a cabo en este momento. Es de hacer notar la importancia que tiene la elaboración de este tipo de pruebas adaptadas a dialectos particulares donde el juego de oposiciones que se presentan puede ser diferente al de otros dialectos del español. Estas diferencias, que podrían no ser tan importantes en el caso de evaluaciones de sistemas de síntesis de voz adquieren relevancia cuando se trata de evaluar la audición humana en casos clínicos.

5. Bibliografía

- P. Hoole. Methodological considerations in the use of electromagnetic articulography in phonetic research. In Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München, volume 31, pages 43-64, 1993.
- W. Marslen-Wilson and P. Warren. Levels of perceptual representation and process in lexical access — words, phonemes, and features. Psychological Review, 101:653-675, 1994.
- Slater and J. Coleman. Non-segmental analysis and synthesis based on a speech database. In Proceedings of the Fourth International Conference on Spoken Language Processing, volume 4, pages 2379-2382, Philadelphia, 3-6 October 1996.
- W. Voiers. Evaluating processed speech using the Diagnostic Rhyme Test. Speech Technology. 1, 30-39. 1983.
- M.R. Cárdenas y V. Marrero. Cuaderno de logoaudiometría. Cuadernos de la UNED. Simancas Ediciones, S.A. Valladolid. 1994.
- E. Obediente. Fonética y Fonología. Universidad de Los Andes. Consejo de Publicaciones. 1998.