

# Psychometric Characteristics of Spanish Monosyllabic, Bisyllabic, and Trisyllabic Words for Use in Word-Recognition Protocols

Mitzarie A. Carlo<sup>1</sup>  Richard H. Wilson<sup>2</sup> Albert Villanueva-Reyes<sup>3,4</sup>

<sup>1</sup>Audiology Program, School of Health Professions, Medical Sciences Campus, University of Puerto Rico, San Juan, Puerto Rico

<sup>2</sup>Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona

<sup>3</sup>School of Health Professions, Medical Sciences Campus, University of Puerto Rico, San Juan, Puerto Rico

<sup>4</sup>Speech-Language Pathology Program, Gannon University, Ruskin, Florida

**Address for correspondence** Mitzarie A. Carlo, AuD, PhD, MSc, Audiology Program EPS 630, School of Health Professions, U.P.R.—Medical Sciences Campus, PO Box 365067, San Juan, Puerto Rico 00936-5067 (e-mail: mitzarie.carlo@upr.edu).

J Am Acad Audiol 2020;31:531–546.

## Abstract

**Background** English materials for speech audiometry are well established. In Spanish, speech-recognition materials are not standardized with monosyllables, bisyllables, and trisyllables used in word-recognition protocols.

**Purpose** This study aimed to establish the psychometric characteristics of common Spanish monosyllabic, bisyllabic, and trisyllabic words for potential use in word-recognition procedures.

**Research Design** Prospective descriptive study.

**Study Sample** Eighteen adult Puerto Ricans ( $M = 25.6$  years) with normal hearing [ $M = 7.8$ -dB hearing level (HL) pure-tone average] were recruited for two experiments.

**Data Collection and Analyses** A digital recording of 575 Spanish words was created (139 monosyllables, 359 bisyllables, and 77 trisyllables), incorporating materials from a variety of Spanish word-recognition lists. Experiment 1 ( $n = 6$ ) used 25 randomly selected words from each of the three syllabic categories to estimate the presentation level ranges needed to obtain recognition performances over the 10 to 90% range. In Experiment 2 ( $n = 12$ ) the 575 words were presented over five 1-hour sessions using presentation levels from 0- to 30-dB HL in 5-dB steps (monosyllables), 0- to 25-dB HL in 5-dB steps (bisyllables), and -3- to 17-dB HL in 4-dB steps (trisyllables). The presentation order of both the words and the presentation levels were randomized for each listener. The functions for each listener and each word were fit with polynomial equations from which the 50% points and slopes at the 50% point were calculated.

**Results** The mean 50% points and slopes at 50% were 8.9-dB HL, 4.0%/dB (monosyllables), 6.9-dB HL, 5.1%/dB (bisyllables), and 1.4-dB HL, 6.3%/dB (trisyllables). The Kruskal–Wallis test with Mann–Whitney U post-hoc analysis indicated that the mean 50% points and slopes at the 50% points of the individual word functions were significantly different among the syllabic categories. Although significant differences were observed among the syllabic categories, substantial overlap was noted in the

## Keywords

- ▶ auditory perception
- ▶ normal hearing for pure tones
- ▶ Spanish word recognition
- ▶ speech perception
- ▶ word recognition

received  
July 2, 2019  
accepted after revision  
December 23, 2019

Copyright © 2020 2020 by the American Academy of Audiology. All rights reserved. Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 760-0888. DOI <https://doi.org/10.1055/s-0040-1709446>. ISSN 1050-0545.

individual word functions, indicating that the psychometric characteristics of the words were not dictated exclusively by the syllabic number. Influences associated with word difficulty, word familiarity, singular and plural form words, phonetic stress patterns, and gender word patterns also were evaluated.

**Conclusion** The main finding was the direct relation between the number of syllables in a word and word-recognition performance. In general, words with more syllables were more easily recognized; there were, however, exceptions. The current data from young adults with normal hearing established the psychometric characteristics of the 575 Spanish words on which the formulation of word lists for both threshold and suprathreshold measures of word-recognition abilities in quiet and in noise and other word-recognition protocols can be based.

English word-recognition materials used in audiology have been investigated extensively, standardized, and even recorded since the 1940s when the principles involved in the word-recognition construct were delineated.<sup>1,2</sup> A multitude of Spanish word-recognition lists have been developed over the years,<sup>3-10</sup> but available, recorded, standardized Spanish materials for audiologic protocols are lacking.

English word-recognition materials, like spondaic words used to establish the speech-recognition threshold (SRT) and monosyllabic words used to evaluate word-recognition performance in a variety of paradigms, have been used as references in the development of Spanish materials for similar tasks. Findings from the English literature must be applied cautiously to Spanish word-recognition materials because of the lexical, phonetic, and phonotactic differences that exist between the two languages. Although Modern English reflects substantial influences from Romance languages, such as Spanish and Italian, the lexicons of each language are distinct. Phonetically, English and Spanish vary in vowel and consonant inventories. For example, Spanish has only five phonetic vowels, whereas English has over 10 (the number of vowels also varies by dialect). Some English phonetic consonants do not occur in Spanish (e.g., /v/, /z/, /ʃ/, and /ɹ/) or only occur in specific Spanish dialects (e.g., /x/ and /θ/). Other phonetic consonants occur in the Spanish language but not in English (e.g., /r/ and /r/).

Phonotactic rules deal with the acceptable combinations of phonemes in a language that define permissible syllable structure, consonant clusters, and vowel sequences. Spanish is more phonotactically constrained than English in the use of consonants.<sup>11</sup> English has over 30 onset consonant clusters that are permissible in the language and over 19 coda clusters.<sup>12</sup> Onset consonant clusters are a combination of two or more consonants before the syllable nucleus, which is often a vowel, whereas coda clusters are consonant combinations following the syllable nucleus. Spanish only allows approximately 12 onset consonant clusters, all of which combine an initial consonant with either /r/ or /l/; the use of /s/ as an initial phoneme in onset clusters is prohibited.<sup>11</sup> In word endings, English allows up to four consonants in a coda cluster, whereas in Spanish coda clusters in word endings are not allowed.<sup>13</sup> Coda consonant clusters in Spanish are allowed in nonfinal positions but only when a

consonant is combined with /s/. Exceptions to Spanish phonotactic constraints, however, may occur with words that have been borrowed from other languages. The strict consonant cluster constraints in Spanish inevitably result in a vowel-dominated language with 71.8% of its syllables ending in a vowel (open syllables<sup>14</sup>). The increased use of vowels in the Spanish language results in a greater frequency of multisyllabic words since syllabification is directly influenced by the number of audible vowels in a word. Likely as a result of these phonotactic constraints, bisyllables are the most frequently occurring words in Spanish followed by trisyllables.<sup>15</sup> With the exception of function words (e.g., articles, pronouns, and prepositions), monosyllables are infrequent in Spanish, which is in contrast with English, in which monosyllabic words constitute up to 71.5% of the vocabulary.<sup>16</sup> Given the aforementioned language differences, it would be inappropriate to attempt to adapt directly English word materials in the development of Spanish speech-recognition tests. General recommendations from the English literature on the development of appropriate word-recognition materials, however, can be followed, such as the use of items that are representative of the language, familiar to the listener, and of average equal difficulty.<sup>2</sup>

Most, if not all, of the word-recognition materials in current use originated with a list of written words that were thought, through one process or another, to be simple and familiar words to the target listeners. Additionally, some linguistic considerations were factored into the selection process. As Hudgins et al<sup>1</sup> and Cancel-Ferrer<sup>4</sup> noted, an equally important variable in the word-recognition process is the intelligibility of the particular utterance of the word under consideration. This utterance-intelligibility factor prompted the initial phase of this project reported here that was to develop a corpus of recorded Spanish words with established psychometric characteristics from which various lists of words can be compiled for speech-perception protocols used in audiological evaluations (e.g., speech threshold and suprathreshold measures, speech-in-noise, dichotic listening, and the variety of distorted speech paradigms). The present study evaluated the recognition performances of young listeners with normal hearing for pure tones on monosyllabic, bisyllabic, and trisyllabic Spanish words recorded by a female speaker and presented at

multiple presentation levels. Two experiments were conducted. Experiment 1 used 25 randomly selected words from each of the three syllabic categories to determine the presentation levels required with each syllabic category to produce word-recognition functions minimally over the 20 to 80% correct range. Experiment 2 generated word-recognition functions on each of the 139 monosyllabic, 359 bisyllabic, and 77 trisyllabic words using a protocol in which both the word order and the presentation levels were randomized. Prior to detailing the two experiments, the material preparations (including word selection, recording, editing, calibrating, and the compilation of the stimulus words for the individual participants) are presented along with general, common characteristics of the participant groups and of the two experimental protocols.

### Material Preparation

A list of 575 Spanish monosyllabic, bisyllabic, and trisyllabic words was created mainly of words belonging to clinically available Spanish speech-recognition lists with some familiar words added. Words common to two or more lists were included only once, a principle that was also applied to homophone words of Spanish dialects that use *seseo* (i.e., *tiza, tisa; lez, les; and taza, tasa*). *Seseo* refers to the pronunciation of the grapheme ⟨z⟩ as /s/, as well as, for the grapheme ⟨c⟩, when encountered before ⟨e⟩ or ⟨i⟩. The 139 monosyllabic words were drawn from Rosenblüt and Cruz<sup>6</sup> and the Auditec of St. Louis *Spanish Monosyllables* (Auditec Item 151; www.auditec.com). Of the 359 bisyllabic words, 344 were from one or more of four sources (*Spanish Words for Speech Recognition Threshold lists*,<sup>4</sup> *Spanish Bisyllables Form 1, Lists A–D* [Auditec, Item 268], the *Spanish Picture Identification Task*,<sup>17</sup> and the Brigham Young University Spanish word-recognition lists<sup>18</sup> with 15 supplemental words added. The added words included two bisyllabic digests that were not included in the source lists and 11 bisyllabic words based on familiarity within the Puerto Rican dialect. Of the 77 trisyllabic words, 36 were from the Auditec *Spanish Trisyllables for SRT* (Auditec, Item 120) with 41 other familiar trisyllabic words added. The 575 words and their sources are listed in Tables S1 to S3 in the Supplementary Material.

### Recordings

Digital recordings of the 575 words were produced in a professional recording studio during five 1.5-hour sessions at the same time each day. Each session included an initial practice period. The recordings were made in an IAC single-wall sound booth with a wind screened microphone (Electro-Voice, Model RE20) routed through a digital mixing console (Yamaha, Model 02R96) to a computer for storage. The bit depth was 16-bits with 44,100 samples/s. The speaker was a professional Puerto Rican female broadcaster with a neutral Spanish accent. During the recording sessions, an audiologist (first author) and a linguist/speech-language pathologist (third author), both native Spanish speakers from Puerto Rico, provided feedback to the speaker on word productions and requested additional utterances when necessary.

The recording sequence was from monosyllabic to trisyllabic words with the order of the words within each syllable type randomized. For each word, the speaker repeated the carrier phrase *diga usted* (say the word), followed by two consecutive utterances of the word, monitoring the level of the carrier phrase on a volume unit (vu) meter.<sup>19,20</sup> This format was replicated a minimum of three times for each word. The interval between the carrier phrase and the first utterance was 200 to 300 ms. In this manner, the carrier phrase and words were independent utterances, making it possible to create future delivery paradigms that were not carrier-phrase dependent, e.g., presenting the target words without a carrier phrase or presenting more than one target word following a carrier phrase. Additional carrier phrases *muéstrame* and *enséñame* (show me) as well as instructions in Spanish for word-recognition tasks also were recorded.

### Audio Editing

A two-step judging process was completed to select the best recorded utterance of each target word. First, the audiologist and the linguist/speech-language pathologist involved in the recording sessions listened to the recordings and selected the best three utterances of each word by qualitatively discussing which were the most natural sounding utterances based on accurate articulation, rate of speech, intonation, and voice projection. The audio file containing only the three best utterances of each word were then listened to by a clinical audiologist (with no previous exposure to the audio recordings) who selected the best utterance of each word based only on auditory perception. The first author was present during the final selection process to assist with the manipulation of the audio but did not provide feedback to the rater. Each word and *diga usted* then were put in separate \*.pcm files, which are \*.wav files without a header. The batch-processing routine in the waveform editor<sup>21</sup> was used to obtain the various presentation levels.

### Word Durations

Descriptive statistics on word duration by syllabic word category are available in Table S4 in the Supplementary Material and can be summarized as follows: monosyllables ( $M = 535$  ms, standard deviation [SD] = 103 ms, max = 750 ms, min = 247 ms), bisyllables ( $M = 611$  ms,  $SD = 95$  ms, max = 953 ms, min = 378 ms), and trisyllables ( $M = 717$  ms,  $SD = 93$  ms, max = 947 ms, min = 571 ms). Thus, there is a marked positive relation between the number of syllables and the duration of the word. The distributions of the word durations for each of the three syllable types are shown in Figure S1 in the Supplementary Material with the individual word durations and English translations provided in Tables S5 to S7. Based on the skewness coefficients shown in Figure S1 and listed in Table S1, the monosyllabic and bisyllabic word durations were approximately symmetrical (–0.5 to 0.5), whereas the trisyllable distribution was moderately skewed (0.5 to 1.0).

### Calibration

The waveforms of a word typically can be characterized as having a large amplitude-modulation component that has a major influence on most overall word amplitude measures like

root-mean-square (rms) and vu levels. In speech audiometry, the amplitudes of materials are set traditionally with a vu meter, which is basically a mechanical average with a time constant of approximately 300 to 350 ms.<sup>22</sup> The intended application of the vu meter was monitoring the signal level in broadcast and transmission line applications, not the precise measurement of signal level (Wilson<sup>23</sup> provides a recent review of the vu meter). Most often the target words have sustained amplitudes whose durations are shorter than the time constant of the vu meter. For these reasons, the vu meter is ineffective in measuring the amplitudes of single words, which is why the ANSI S3.16<sup>24</sup> standard specifies monitoring the level of the carrier phrase followed by the naturally spoken target word. The vu level of the carrier phrase should correspond to the vu level of the 1,000-Hz calibration tone.

As the target words were uttered independent of a carrier phrase, an rms technique was needed to “equate” the levels of the words. Almost exclusively, the largest signal amplitudes in words are achieved by a segment of the vowel, which compared with the consonants is fairly sustained. The majority of the 575 words were multisyllables with most having one vowel whose amplitude was greater than the other vowel(s) in the word. As illustrated in ►Fig. 1, a visual determination was made of the maximum-amplitude segment of the word and the rms of that segment was calculated. Note from the figure that the durations of the maximum-amplitude segments varied substantially and one word, *tres*, had a short-lived maximum amplitude in a nonvowel segment /r/. The overall amplitude of each word then was adjusted so that the

amplitude of the maximum amplitude segment of each word was set to the same rms. The carrier phrase was also adjusted to the same rms as the target words. This calibration technique was not intended to produce equal intelligibility among words but was intended to provide a replicable calibration technique that can be applied to relatively short signals (note: for all words in a more recent study, the maximum amplitude of the vowel has been defined as the 50-ms segment of the vowel with the maximum amplitude).<sup>25</sup>

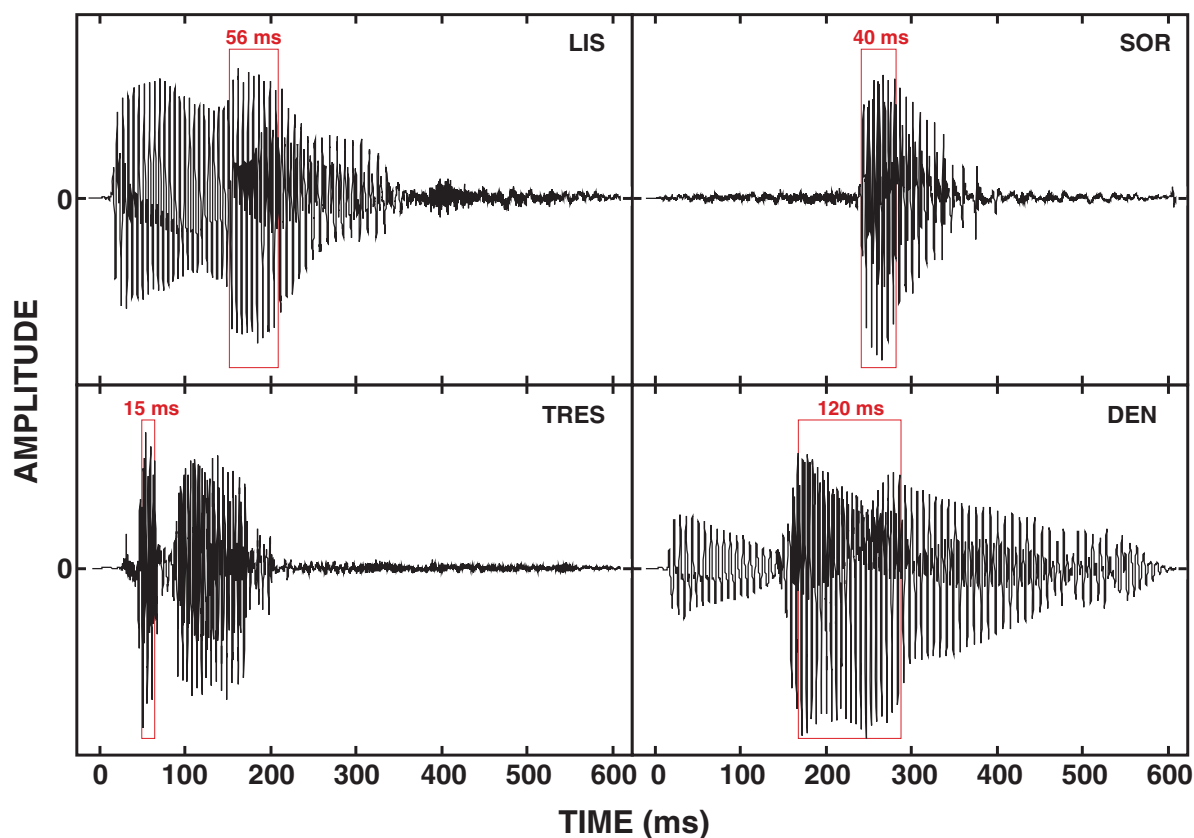
## General Methods

### Participants

All participants were Puerto Ricans from the University of Puerto Rico, Medical Sciences Campus and nearby communities who learned Spanish as a first language and used Spanish as their main mode of communication. The experiments were approved by the Institutional Review Board of the University and the investigation was conducted with the understanding and full, informed consent of the subjects.

### Procedures

The recordings were converted from \*.wav files to Apple Lossless Audio Codec (APAC) files using Apple Inc. iTunes Software (version 12.1) and uploaded to a portable media player (Apple Inc., iPod Touch, Cupertino, CA) that was fed to an audiometer (Grason-Stadler, Model 61; Eden Prairie, MN) and presented monaurally through a TDH-49 earphone (Telephonics, Farmingdale, NY) encased in a supra-aural



**Fig. 1** Waveforms of four words highlighting the maximum amplitude segments (ms) that were utilized for the rms calculations.

cushion. The 1-hour test sessions, which included a break midway through the session, were conducted in a double-walled booth (Industrial Acoustics Company, Model 120act-3, North Aurora, IL). Prior to the experimental stimuli, the participants listened to recorded instructions that included two example stimuli mimicking the word-recognition task. The instructions requested that the participants repeat each word they heard and to guess when unsure of what they heard. The responses to each word were recorded into a spread sheet. For both experiments, the participants were compensated \$10 for each completed session.

## Experiment 1

The goal of Experiment 1 was to provide guideline data on the presentation levels required to produce psychometric functions for the monosyllabic, bisyllabic, and trisyllabic Spanish words minimally over a 60% recognition-performance range from 20 to 80% with ideally an equal number of datum points below and above 50% correct. The guideline data were established at seven presentation levels with 25 words randomly selected from each syllabic category.

### Participants

Two male and four female young listeners ( $M = 28.0$  years,  $SD = 5.5$  years) with normal hearing at the 250 to 8,000 Hz octaves ( $\leq 20$ -dB hearing level [HL]<sup>24</sup>) participated. The mean thresholds ranged from 2.5- to 10.8-dB HL with a three-frequency (500, 1,000, and 2,000 Hz) pure-tone average (PTA) of 8.3-dB HL ( $SD = 4.3$  dB). Right ears were tested on half of the listeners.

### Procedures

The participants completed a single session in which 525 stimuli (25 words  $\times$  3 syllabic conditions  $\times$  7 presentation levels) were presented randomly. The seven presentation levels ranged from -5- to 25-dB HL in 5-dB steps. A 150 ms silent interval separated the carrier phrases and the target words with a 3-s interstimulus interval (ISI). For each participant, a randomization of the 525 word files was created and then presented via 21 tracks of 25-words each. A schematized 25-word example track is depicted in Figure S2 in the Supplementary Material.

### Results

The percent correct recognition performances for each syllabic group at each presentation level were computed to develop mean recognition functions for the monosyllabic, bisyllabic, and trisyllabic word groups and are presented in **Fig. 2** (upper left panel); the data in the three remaining panels are discussed with Experiment 2. Third-degree polynomials were used to describe the data. The recognition data for the 25 words of each syllabic groups used in Experiment 1 are listed in Tables S8 to S10. The obvious observation is the direct relation between recognition performance and number of syllables in the words. As the number of syllables increased, recognition performance increased. As reminded by Egan<sup>2</sup> (page 961), the 50% point on a psychometric function is the most sensitive

region of a function because it is furthest from floor and ceiling effects. The polynomial equations were used to calculate the hearing level (dB) of the 50% points and the first derivatives of the polynomials were used to calculate the slopes of the functions (%/dB) at the 50% points. For the monosyllabic, bisyllabic, and trisyllabic words, the 50% points were 13.3-, 8.4-, and 2.9-dB HL, respectively, with corresponding slopes of the mean functions at the 50% points of 3.9, 4.2, and 5.2%/dB. The maximum recognition performances ( $>90\%$ ) were not obtained by the monosyllabic words at 25-dB HL but were achieved with the bisyllabic words at 25-dB HL and with the trisyllabic words at 20-dB HL. Minimum performances ( $<10\%$ ) were around 0-dB HL for the monosyllabic and bisyllabic words and extended below 0-dB HL for the trisyllabic words. These preliminary data provided the presentation-level guidelines most appropriate for the subsequent experiment to obtain datum points throughout the ranges of recognition performances for each of the three syllabic categories.

## Experiment 2

The objective of this experiment was to determine the psychometric characteristics of each of the 575 words that can be used in the future development of Spanish word-based protocols for use in the variety of audiologic clinical and research applications.

### Participants

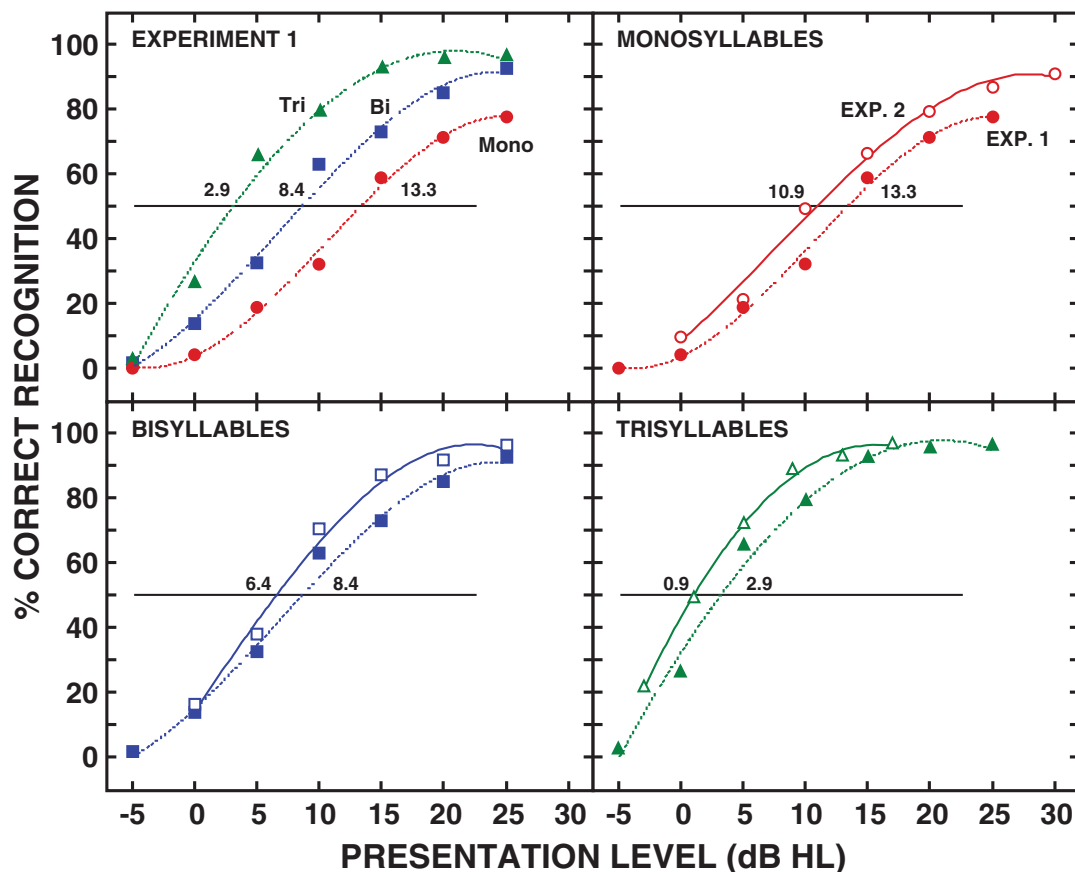
Three male and nine female young adults with normal hearing at the 250 to 8,000 Hz octaves ( $\leq 20$ -dB HL) and between 21 to 33 years of age ( $M = 24.3$  years,  $SD = 4.7$  years) participated. The mean PTA was 7.5-dB HL ( $SD = 2.5$  dB). Seven of the participants were tested in the right ear. An educational inclusion criterion of no greater than an undergraduate degree was added in an effort to maintain the study sample representative of the young-adult Puerto Rican population.

### Procedures

A separate audio file was created for each of the 973 monosyllabic stimuli (139 words  $\times$  7 presentation levels), each of the 2,154 bisyllabic stimuli (359  $\times$  6 presentation levels), and each of the 462 trisyllabic stimuli (77 words  $\times$  6 presentation levels). Based on the data from Experiment 1, the following presentation levels were used: (1) monosyllables, 0- to 30-dB HL in 5-dB increments, (2) bisyllables, 0- to 25-dB HL in 5-dB increments, and (3) trisyllables, -3 to 17-dB HL in 4-dB increments. The 3,589 word files were randomized for each participant and then grouped into 143, 25-word lists plus one list of 14 words, which were parsed into the five 1-hour sessions. The interval each listener took to complete the five test sessions varied from 6 to 63 days ( $M = 26.6$  days,  $SD = 16.4$  days). The silent interval between the carrier phrase and the target word was 200 ms and the ISI was 2 s.

### Results and Discussion

In this section the stability of the recognition performances across the five test sessions are presented followed by various central tendency measures of the monosyllabic,



**Fig. 2** Recognition-performance functions for the randomly selected 25 monosyllabic words (red circles), the 25 bisyllabic words (blue squares), and 25 trisyllabic words (green triangles) obtained in Experiment 1 (filled symbols) from six young listeners with normal hearing are shown in the upper left panel. For comparison, the functions for the same 75 words obtained in Experiment 2 (open symbols) from 12 young listeners are included in the remaining three panels. Third-degree polynomials are used to describe the data. The numbers in each panel are the hearing levels (dB) at which the 50% points occurred. The individual listener data for Experiment 1 are listed in Tables S5 to S7 in the Supplementary Material.

bisyllabic, and trisyllabic word data. Then, the recognition performances are considered for the individual listeners, followed by the individual words, which were analyzed on the performance continua from the easiest to the hardest to understand. A further focus is presented with the words in the following unique categorings: (1) familiar versus unfamiliar words, (2) singular versus plural forms of the same words, (3) last syllable versus penultimate syllable stress patterns, and (4) female and male gender words. Finally, a brief comparison is made of the current data with similar data from earlier investigations.

The mean overall percent correct recognition performances for sessions 1 to 5 were 64.0, 64.5, 64.4, 67.1, and 66.8, respectively. There was good consistency across the five test sessions with, as expected, a hint of improved performance at the 50% point of <1 dB across sessions. The mean performance data across sessions for the individual listeners for each of the three syllable types are shown in Figure S3 that also includes the linear regressions used to describe the mean data. The slopes of the regressions, which ranged from 0.3%/session (monosyllables) to 1.3%/session (trisyllables), emphasize the minimal effect that session had on overall recognition performance.

The mean recognition performances (and standard deviations) for the three syllabic categories are listed in ►Table 1

and depicted in ►Fig. 3 (upper left panel), in which third-degree polynomials are used to describe the data. The 50% points for the mean functions calculated with the polynomials for the monosyllabic, bisyllabic, and trisyllabic words were 8.9-, 6.9-, and 1.4-dB HL, respectively, with corresponding slopes at the 50% points of the mean functions (calculated with the first derivative of the polynomial equations) of 4.0, 5.1, and 6.3%/dB. The more traditional linear slopes of the functions ( $m = \Delta y / \Delta x$ ) between the 20 and 80% points (linear<sub>20-80%</sub>) were 0.2 to 0.3%/dB less steep, mainly because the functions were not linear throughout much of the range of presentation levels. The three mean functions were steepest at the lower presentation levels and progressively became more gradual as the presentation level increased. The functions were linear only over a short segment of the bisyllabic word function from approximately 2- to 5-dB HL. (Note: the slope functions of the mean data from the polynomial first derivatives are plotted in Figure S4.)

An indication of the test, retest characteristics of the syllabic utterances is provided by comparing recognition performances on the 75 words in Experiment 1 with the performances on the same 75 words in Experiment 2 (25 words in each syllabic category), which are illustrated in ►Fig. 2. Visual inspection of the functions in ►Fig. 2 indicate

**Table 1** Mean percent correct recognition and standard deviation by presentation level and 50% recognition point obtained from the third-degree polynomial equation for the monosyllabic, bisyllabic, and trisyllabic words

dB HL	Monosyllables		Bisyllables		Trisyllables	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
30	92.1	12.9				
25	89.1	16.2	95.9	8.9		
20	84.2	17.7	93.4	11.8		
17					97.3	5.1
15	74.3	23.1	86.2	17.9		
13					93.9	9.7
10	55.8	24.9	67.2	24.2		
9					88.2	14.0
5	29.6	20.5	37.4	25.5	72.4	21.0
1					45.6	23.0
0	11.9	12.1	14.5	16.1		
-3					20.1	15.0
50% point (dB HL)	8.9		6.9		1.4	
Slope of the mean at 50% (%/dB)	4.0		5.1		6.3	
Mean of the slopes at 50% (%/dB)	5.1	1.7	6.8	1.9	7.9	2.6

Note: Slopes at the 50% point on the polynomial function (function of the mean) and averaged from the individual 50% points (mean of functions) also are included.

similar recognition functions from the two experiments with performances from Experiment 2 better throughout the presentation-level ranges than the performances from Experiment 1. At the 50% points, there was a 2.4-dB difference between the functions for the monosyllabic words and 2.0-dB differences for both the bisyllabic and trisyllabic words. The slopes of the functions at the 50% points were the same for monosyllabic words (3.9 and 4.0%/dB) and were approximately 1.0%/dB steeper in Experiment 2 for the bisyllables (4.2 and 5.1%/dB) and trisyllables (5.2 and 6.3%/dB). The slight differences between recognition performances in the two experiments are reasonable and can be attributed to listener group differences and the differences in protocols, *viz.*, more exposure to the listening task in Experiment 2.

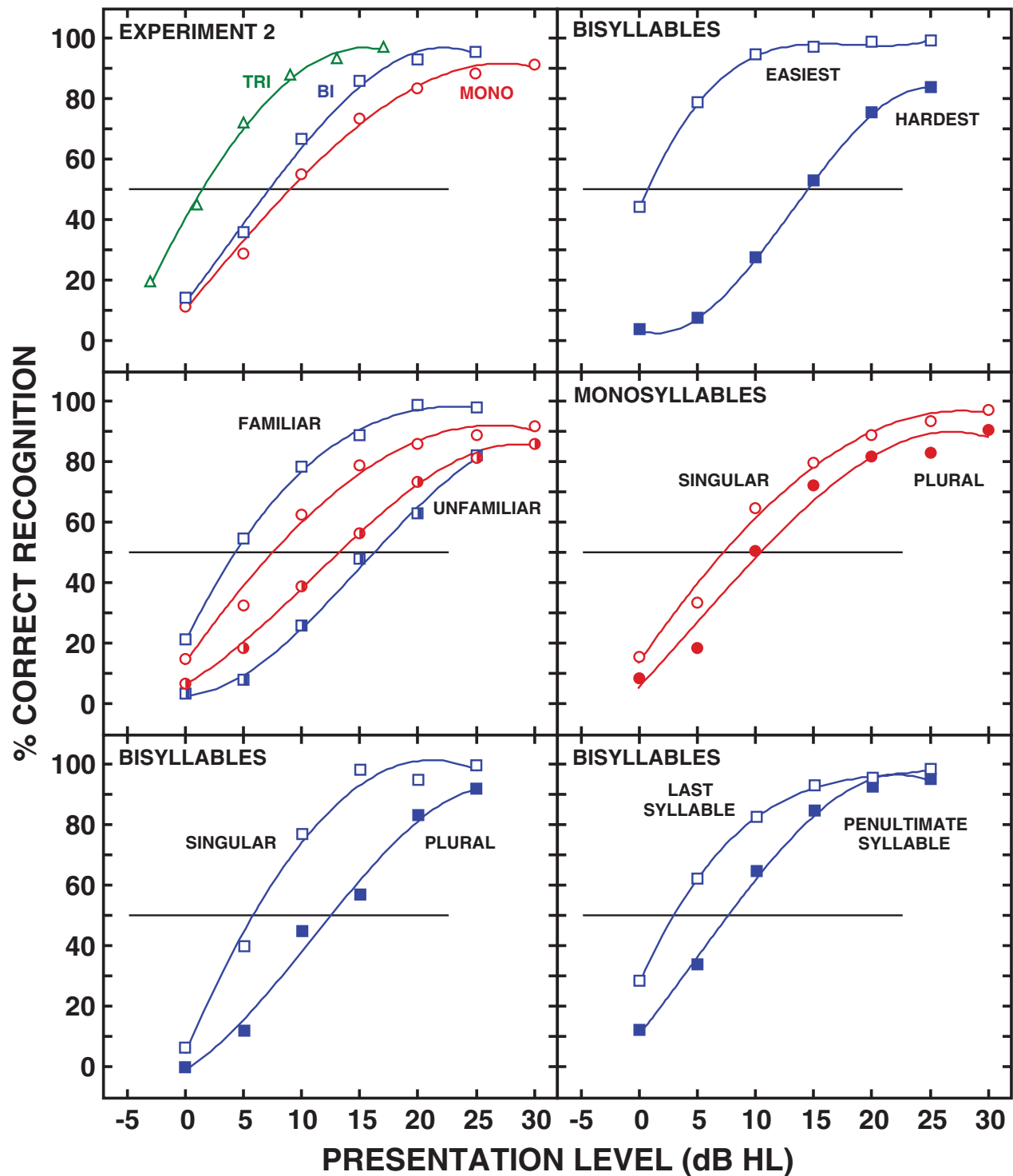
#### Individual-Listener Recognition Performances

Examination of the recognition performances by the individual listeners provides insights into the data not realized from the various measures of central tendency. A major example of these insights involves the slopes of the mean functions presented a couple of paragraphs earlier. As Wilson and Margolis<sup>26</sup> demonstrated, the mean of the individual slopes provides a better estimate of the true slope of a function than does the slope of the mean function. Additionally, the intersubject variability of the measure is available from the individual data. Based on the polynomials used to describe the individual subject data illustrated in ▶Fig. 4, the mean 50% points (and standard deviations) for the monosyllables, bisyllables, and trisyllables were 8.8-dB HL (*SD* = 2.7 dB), 6.9-dB HL (*SD* = 2.6 dB), and 1.3-dB HL (*SD* = 2.7 dB), respectively, with the corresponding slopes at the 50% points of 4.3%/dB (*SD* = 0.4%/dB), 5.5%/dB (*SD* = 0.5%/dB),

and 7.3%/dB (*SD* = 1.0%/dB). Thus, the 50% points from the individual functions and from mean functions presented earlier are essentially identical, whereas the slopes are slightly steeper when calculated from the individual-listener data. (Note: the raw data from each listener in Experiment 2 are listed in Tables S11–S13 with the data calculated from the polynomials used to describe the functions listed in Tables S14–S16.)

The recognition functions for the three syllabic categories depicted in ▶Fig. 4 for each of the 12 young listeners demonstrate the diversity of relations between the various functions. Except for Subject 1, the functions are systematic with the trisyllabic words the easiest to understand and the monosyllabic words the most difficult. The following are measures made at the 50% points on the recognition functions. For Subject 1, the bisyllabic and monosyllabic functions are essentially the same with recognition performance on the monosyllabic words 0.6 dB better than on the bisyllabic words. For the monosyllables, S5 had the best performance (5.4-dB HL) and S6 had the poorest performance (12.9-dB HL). With the bisyllables, S5 also had the best performance (3.0-dB HL) and S6 the poorest (10.8-dB HL). For the trisyllables, S9 had the best performance (-2.2-dB HL) and S3 had the poorest (5.2-dB HL). Finally, the slopes of the individual subject functions ranged from 3.6%/dB (S11) to 5.0%/dB (S1) for the monosyllables, from 4.6%/dB (S11) to 6.5%/dB (S3) for the bisyllables, and from 5.6%/dB (S4) to 8.6%/dB (S11) for the trisyllables. These analyses demonstrate the more in-depth examination that can be gleaned from the individual data compared with the mean data.

The recognition functions in ▶Fig. 4, which are from a relatively homogeneous group of young listeners,

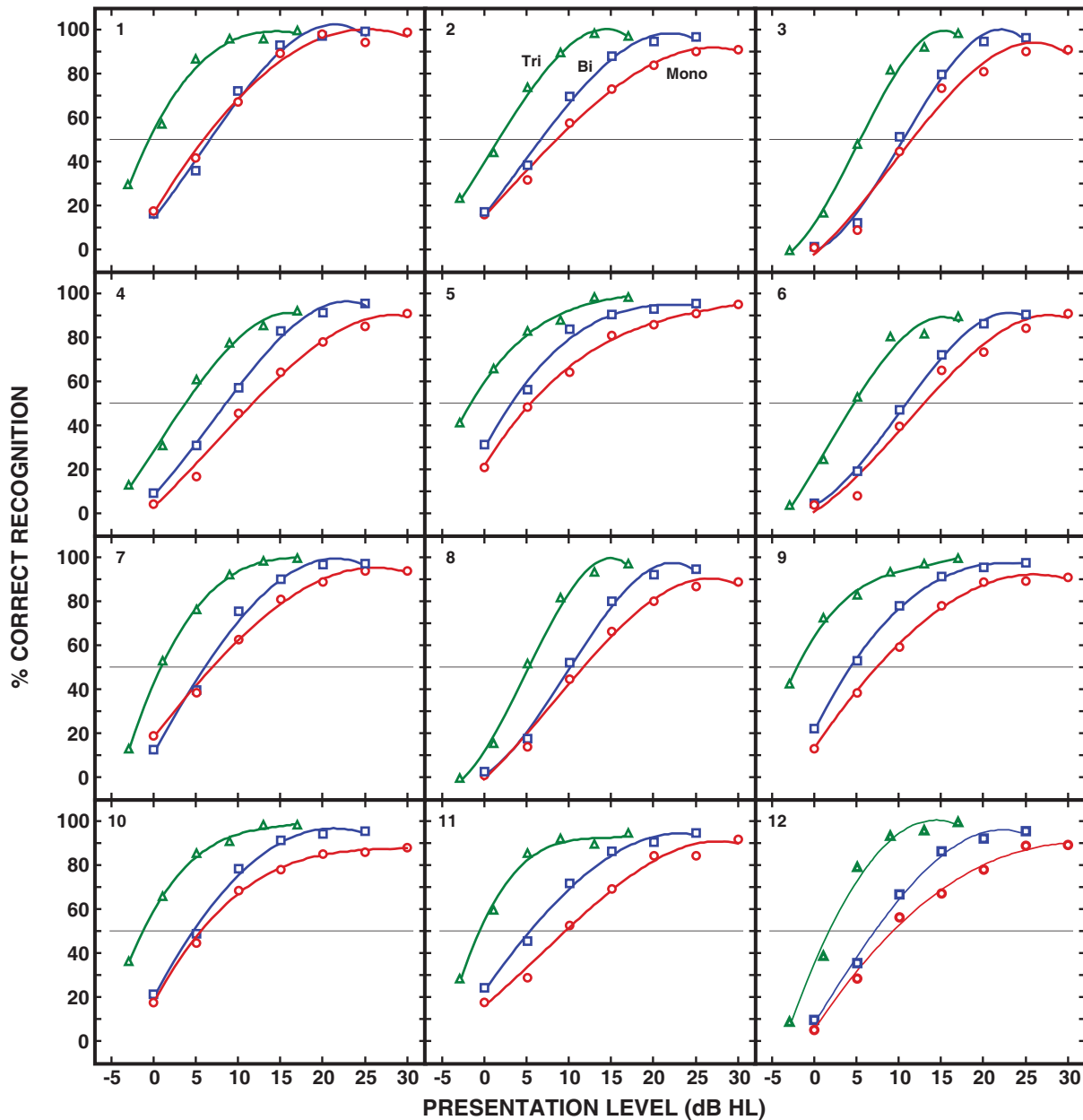


**Fig. 3** Upper left panel: the mean recognition-performance functions for the 139 monosyllabic words (circles), 359 bisyllabic words (squares), and 77 trisyllabic words (triangles) obtained in Experiment 2 from 12 young listeners are displayed along with the third-degree polynomials used to describe the data. Upper right panel: the mean recognition functions of the 50 easiest and 50 hardest bisyllabic words to understand. Middle left panel: the mean recognition functions from 25 monosyllabic and 9 bisyllabic words categorized as unfamiliar are depicted (partially shaded symbols) with a like number of randomly selected words (open symbols). Middle right panel: the mean recognition functions for the singular and plural forms of 15 monosyllabic words. Lower left panel: the functions for the singular and plural forms of 5 bisyllabic words. Lower right panel: the mean recognition functions for 41 words bisyllabic words with phonetic stress on the last syllable and of 318 bisyllabic words with phonetic stress on the penultimate syllable.

demonstrate that across subjects the patterns of responses for the three syllabic categories are similar but within the data of each listener, the interfunction relations are varied. Finally, an interesting observation from these data are that all

36 functions (12 listeners  $\times$  3 syllable lengths) were about twice as steep at 20% correct compared with 80% correct. Specifically, on average the slopes of the functions were 2.4 to 4.4%/dB steeper at the 20% points than at the 80% points.





**Fig. 4** The mean recognition-performance functions for the 139 monosyllabic words (red circles), the 359 bisyllabic words (blue squares), and 77 trisyllabic words (green triangles) obtained from each of the 12 listeners in Experiment 2 are displayed along with the third-degree polynomials used to describe the data. The individual listener data are listed in Tables S8 to S10 in the Supplementary Material.

For the monosyllabic, bisyllabic, and trisyllabic words, the slopes at 20% correct were 5.0, 6.0, and 8.9%/dB, respectively, whereas at 80% correct the corresponding slopes were 2.6, 3.8, and 4.5%/dB. This slope asymmetry between the 20 and 80% points was observed with most of the recognition functions examined in this study and was highlighted in the slopes of the mean functions illustrated in Figure S4.

#### Individual Words with the Extremes of the Recognition Performance Continua

The raw data for the individual words are listed in Tables S17 to S19, along with the computed 50% points and slopes for

each. Twenty-one of the individual words (3 monosyllables, 17 bisyllables, and 1 trisyllable) had 50% points outside of the presentation levels used in the study and were excluded from the analyses of the 50% points and the slopes of the functions at the 50% points. For the individual words, the mean 50% points and slopes at 50% for the monosyllabic, bisyllabic, and trisyllabic words were 9.5-dB HL ( $SD = 4.9$  dB, 5.5%/dB), 7.4-dB HL ( $SD = 4.5$  dB, 6.8%/dB), and 1.8-dB HL ( $SD = 3.4$  dB, 8.0%/dB), respectively (note: the 50% points and the slopes at the 50% points established with the individual word functions [ $n = 200$ ] and with the individual listener functions [ $n = 12$ ], although based on the same raw data, are slightly

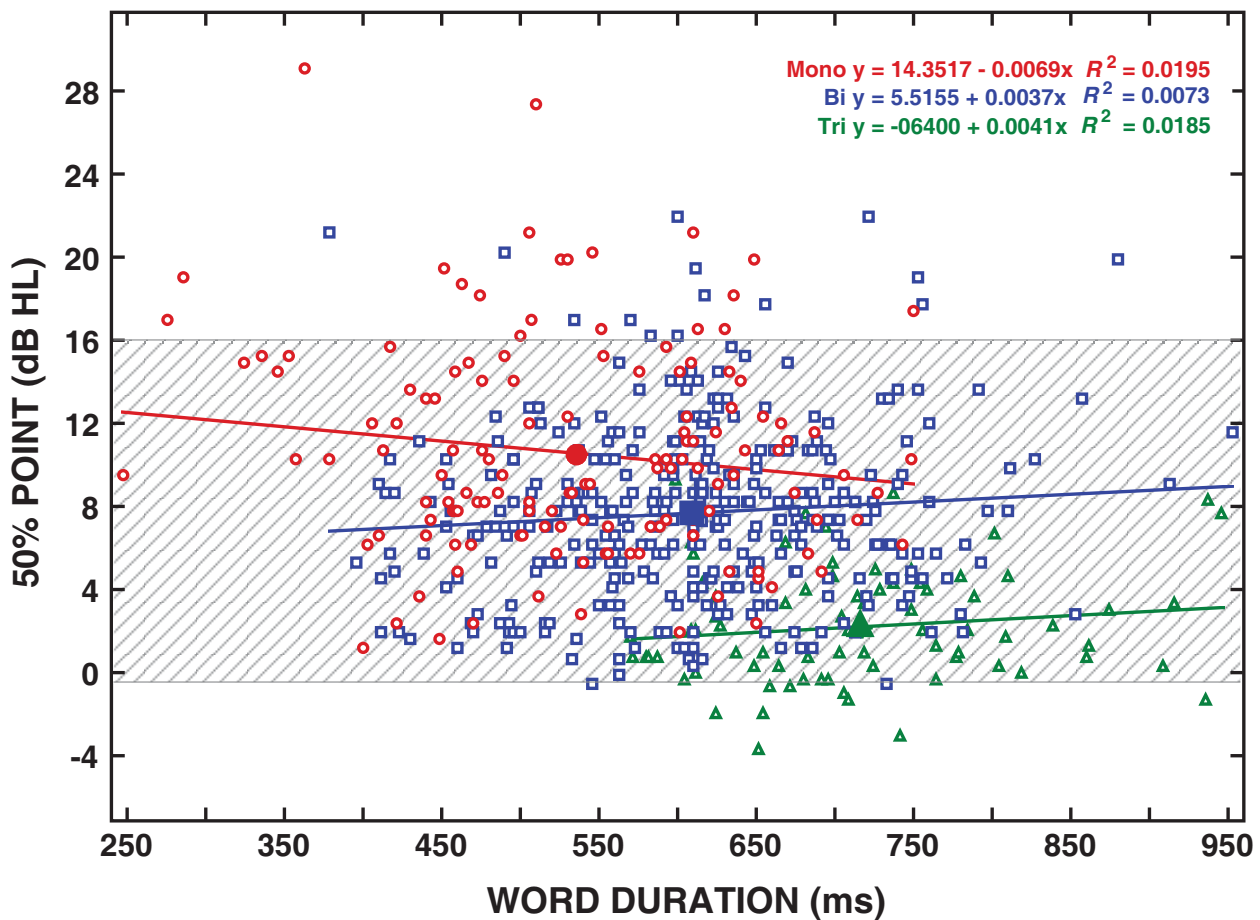
different because of differences in the number of samples and the characteristics of the polynomials used to describe the respective groupings of raw data.). The frequency distributions of the 50% points of the individual words are displayed in Figure S5. The Shapiro–Wilk test revealed that only the trisyllabic words were normally distributed ( $p > 0.05$ ). Nonparametric statistical tests were utilized to compare the individual word data. The Kruskal–Wallis test revealed a significant effect of the syllabic category on both polynomial 50% points [ $\chi^2(2) = 125.1, p < 0.001$ ] and slopes at the 50% points [ $\chi^2(2) = 108.7, p < 0.001$ ]. The Mann–Whitney U post-hoc test with Bonferroni correction showed that the three syllabic categories were significantly different from each other for both the 50% points and the slopes at the 50% points ( $p < 0.001$ ). For comparison, the Spearman–Kärber 50% points for the individual words also are included in Tables S17 to S19.<sup>27</sup> The mean differences between the word 50% points calculated with the polynomial and Spearman–Kärber equations were  $< 1$  dB ( $SD < 2$  dB) for the three syllabic categories indicating that the two measures of the 50% points produce very similar results.

The range of variability for the 50% points calculated from the polynomials for the individual words were from 0.2- to 22.5-dB HL for the monosyllables, from -2.7- to 22.8-dB HL for the bisyllables, and from -6.4- to 12.6-dB HL for the trisyllables. The extent of this variability is illustrated in Figure S6 with the recognition functions from 50 randomly selected words from each syllabic category. The large ranges of 50% points, which was  $> 20$  dB for the three syllable categories, were of interest and were explored in more detail initially with the 50 bisyllabic words with the best overall recognition performances and the 50 bisyllabic words with the poorest overall performances; the use of 50 words (13.9% of 359 words) was arbitrary. Subsequently, the analysis was extended to 19 monosyllabic words and 11 trisyllabic words at both ends of the performance continuum, which was in proportion (13.9%) to the 50 bisyllabic words. The data for the individual words in these groupings are listed in Tables S20 to S24.

The mean recognition functions for the easiest and hardest bisyllabic words are illustrated in ►Fig. 3 (upper right panel). The displacement of the two functions throughout the range of presentation levels is obvious with a 13.8-dB difference at the 50% point (14.4-dB HL minus 0.6-dB HL), which closely mirrored the 13.1-dB difference calculated with the Spearman–Kärber equation (14.9-dB HL minus 1.8-dB HL). The slope of the easier function at 50% was 3.4%/dB steeper than the slope of the harder function (8.6 vs. 5.2%/dB). Similar relations were observed between the mean recognition functions of the easiest and hardest monosyllabic and trisyllabic words, which are depicted in Figure S7. Some of the extreme individual monosyllabic word-performance outliers are illustrated in Figure S8. One recognition function in Figure S8 is irregular (*pon*) because recognition performance increased systematically between 0- and 20-dB HL with a substantial decrease in performances at 25- and 30-dB HL. The outliers represented by *hay*, *tos*, and *ved* can only be considered outliers for these particular utterances of the words.

The morphologies of the functions for the easiest words and for the hardest words were both different and the same within and between the three syllabic categories. Within each of the three syllabic types, the functions for the easiest words were displaced to the lower presentation levels and about twice as steep as the functions for the hardest words. The three syllabic functions for the easiest words were identical in shape; likewise, the three functions for the hardest words were almost identical with only the slopes noticeably different. As the number of syllables increased, the two functions within the syllabic categories both moved to lower presentation levels and came closer together. Between the monosyllabic and trisyllabic words, the 50% points differed (1) by 5 dB from 2.6-dB HL (monosyllables) to -2.4-dB HL (trisyllables) for the easiest words and (2) by 13.2 dB from 19.9-dB HL (monosyllables) to 6.7-dB HL (trisyllables) for the hardest words. The slopes of the functions at the 50% points changed (1) 3.6%/dB from 7.2%/dB (monosyllables) to 10.8%/dB (trisyllabic) for the easiest words and (2) 2.4%/dB from 2.9%/dB (monosyllables) to 5.3%/dB (bisyllables and trisyllables) for the hardest words. The differences between the functions for the easiest and hardest words within each syllabic type at the 50% point (Figure S7) ranged from 17.3 dB (monosyllables) to 9.1 dB (trisyllables). Thus, as syllabic length increases, recognition performances at the 50% points on the so-called “easy” words (intelligibility-wise) demonstrated minimal change with the most noticeable performance changes observed with the “hard” words.

With the individual words, although the mean data indicated a direct relation between syllabic length and recognition performance, it was of interest to explore within each syllabic category the relation between word duration and recognition performance. For these analyses, the bivariate plot shown in ►Fig. 5 was developed using for each word the duration discussed earlier (abscissa) and the 50% point derived with the Spearman–Kärber equation (ordinate). Linear regressions then were generated for each of the three syllabic categories. The results of these analyses demonstrate basically flat regressions with slopes of -0.0069 dB/ms (monosyllables), 0.0037 dB/ms (bisyllables), and 0.0041 dB/ms (trisyllables), all of which indicate a random or no relation between the word duration and recognition performance. Another observation from the data in this format is the intermingling of many of the monosyllabic, bisyllabic, and trisyllabic word 50% points. The extent of the overlap in recognition performances among the words in the three syllabic categories was examined using  $\pm 2$  SDs about the mean 50% recognition performance on the bisyllabic words as the benchmark to quantify overlapping performances. This  $\pm 2$  SD range for the bisyllables (Table S18) was 16.4 dB from -0.4- to 16.0-dB HL, which is depicted in ►Fig. 5 as the shaded area. The algorithm was simple, how many monosyllabic and trisyllabic words had 50% points that were within this 16.4-dB range. Of the 139 monosyllabic words, 124 (89.2%) also had their 50% recognition performances within this  $\pm 2$  SD range of the bisyllabic words, as did 65 (82.3%) of the 77 trisyllabic words. Even using the  $\pm 1$  SD range from the bisyllables (8.2 dB), 61.2 and 32.9% of the monosyllables and trisyllables, respectively, had 50% points



**Fig. 5** The presentation levels (dB HL) at which the 50% points calculated with the Spearman–Kärber equation occurred as a function of the word durations (ms) for the 139 monosyllabic words (red circles), the 359 bisyllabic words (blue squares), and 77 trisyllabic words (green triangles) obtained from the 12 listeners in Experiment 2 are displayed. Linear regressions are used to describe each set of data. The shaded area represents  $\pm 2$  standard deviations of the recognition performances on the bisyllabic words. The numeric data are listed in Tables S2–S4 and S8–S10 in the Supplementary Material.

within that reduced range of performances. Collectively, these data on the individual words attest to the inherent variability associated with interword recognition performances and how mean data used in isolation obscure the underlying volatility that typifies word-recognition tasks.

In summary, the above results delineate recognition performances on word utterances using a variety of parameters. Perhaps the most straightforward comparison of performances is the average or overall performance of each word across the range presentation levels (overall % in Tables S17–S19). Here, the interest is in words that are difficult to understand, even for young adults with normal hearing sensitivity. Words with an average performance  $<40\%$  correct across the various presentation levels arbitrarily were considered difficult words for intelligibility purposes. Of the 139 monosyllabic words, 13 had overall performances  $<40\%$  correct, two of which were  $<30\%$ . Of the 359 bisyllabic words, 15 had overall performances  $<40\%$ , seven of which were  $<30\%$ . None of the 77 trisyllabic words had overall performances  $<40\%$ . These 28 words should be excluded from future lists intended for use with individuals having sensorineural hearing loss. These difficult words refer to the particular utterances of the words utilized in the study and

should not necessarily be generalized to other utterances of the same words, even by the same speaker.

#### Familiar and Unfamiliar Words

As mentioned earlier, the use of familiar or common words is critical in achieving an effective word-recognition instrument.<sup>2,28,29</sup> The vast majority of words compiled for this study were taken from word-recognition lists that originated in a variety of Spanish-speaking countries. This diversity of sources possibly included words that were not familiar to the typical Puerto Rican. Although the evaluation of the effect of word frequency of occurrence on word recognition was beyond the scope of the current study, a brief *a posteriori* survey of word familiarity was conducted to explore the possible confounding effect of word familiarity on recognition performances. The collaborating linguist identified 27 monosyllables and 14 bisyllables from the 575-word list that were not commonly used in Puerto Rico and could be considered locally unfamiliar. Defining familiarity as the ability to define or assimilate the meaning of the word, 37 undergraduate students (25 females and 7 males) were given the 41 words and instructed to write a definition or an associated term for each of the words. The survey revealed

that 25 of the 27 monosyllabic words (92.6%) and 9 of the 14 bisyllabic words (64.3%) surveyed were unfamiliar to at least 75% of the participants; the survey results are included in Table S25. The recognition performances on these unfamiliar words are displayed in ►Fig. 3 (middle left panel) for the 25 monosyllabic words (half-filled circles) and the nine bisyllabic words (half-filled squares). The unfilled symbols are the recognition functions for 25 monosyllables and nine bisyllables randomly selected familiar words included for comparison (note: because only nine words were in the bisyllabic group, six iterations of the random selections were conducted to confirm that the familiar bisyllabic recognition function in ►Fig. 3 was representative). The 50% points on the functions for the familiar and unfamiliar words, respectively, were 7.5- and 13.1-dB HL (monosyllables) and 4.2- and 16.0-dB HL (bisyllables) with the slopes of the familiar functions at 50% approximately 1%/dB steeper. Thus, the unfamiliar words required 5.6 to 11.8 dB higher presentation levels to achieve 50% correct than did the familiar words and the slopes of the unfamiliar word functions were more gradual than the slopes of the familiar word functions. These differences in performance are consistent with the relation between word familiarity and word-recognition performance acknowledged in the Neighborhood Activation Model,<sup>30</sup> which proposes that spoken-word recognition is influenced by both bottom-up and top-down processes.

#### Singular- and Plural-Form Words

From the pool of 575 words, 24 monosyllables and 15 bisyllables are either plural forms of a word, conjugated verbs, or prepositions that end in /n/ or /s/ and retain meaning when the final phoneme is removed (e.g., *ranas*, *ven*, *sin*). An error analysis of these words was conducted to establish the percent of recognition errors owing to (1) the omission of the final phoneme /n/ or /s/, (2) recognizing the word as a totally different word, and (3) the participants not providing a response. For the 24 monosyllabic words 14, 54, and 32% of the errors were attributable to the three types of errors, respectively, as were 39, 33, and 28% of the errors with the 15 bisyllabic words. This analysis prompted a further examination of the possible effects that the use of plural forms might have on word-recognition performance.

Fifteen of the 24 monosyllabic and 5 of the 15 bisyllabic plural words had a singular form or conjugation companion word within the 575-word corpus. Possible pairs in which word meaning changed upon deletion of final /s/ or /n/ (e.g., *me*, *mes*) or in which there was not an exact word match (e.g., *cesta*, *cestos*) were excluded from the analysis to control for word familiarity. The mean data comparing forms are shown in ►Fig. 3 for the monosyllabic words (middle right panel) and bisyllabic words (lower left panel); the data for the individual words are depicted in Figures S9 and S10, respectively. In the case of the verbs *da*, *de*, *va*, and *ve*, both the conjugations with final /s/ and final /n/ were available for comparison and are presented in Figure S9. Nine of the 15 monosyllabic word pairs had 50% points that were 1.7 to 10.0-dB lower on the singular form word than on the plural form word. Four of the monosyllabic pairs had 50% points 1.1 to 6.3-dB higher on the singular

form, and two of the compared pairs had 50% points that differed by <1 dB. The mean data shown in ►Fig. 3 for the 15 singular and plural monosyllabic word pairs demonstrated a 3.2-dB difference at the 50% point with the singular form lower (7.2-dB HL) than the plural form (10.4-dB HL). This approximately 3-dB difference for the monosyllabic words was maintained throughout the range of presentation levels. The slopes of the functions were only minimally different, 4.4%/dB (singular) and 4.1%/dB (plural). Each of the five bisyllabic word pairs, depicted in Figure S10, demonstrated better recognition performances with the singular form than with the plural form with differences at the 50% points ranging from 2.7 dB (*libro*, *libros*) to 11.6 dB (*cama*, *camas*). The mean 50% point (and slope) for the five bisyllabic singular and plural words shown in ►Fig. 3 were 5.7-dB HL (6.6%/dB) and 12.4-dB HL (4.8%/dB), respectively. The recognition error analysis was repeated by looking only at the plural and singular word pairs to determine if the addition of the final phoneme was as common as the deletion of the final phoneme in these cases. The recognition errors for the plural forms of the word pairs (17% for monosyllables and 43% for the bisyllables) were very similar to the values previously discussed for plural final-phoneme deletions (14% for monosyllables and 39% for the bisyllables). In contrast, the recognition errors for the addition of a final /s/ or /n/ to the singular forms of the words only accounted for 8% of the monosyllabic errors and <1% of bisyllabic errors. Error analysis was also performed on eight bisyllabic words that also have a final phoneme of /n/ or /s/ but that do not retain meaning if the final phoneme is omitted (e.g., *lunes*, *joven*). Interestingly, only 1% of recognition errors for these words were due to deletion of the final phoneme, which is in great contrast with the 39% of these errors in the plural bisyllabic forms.

The following general tendencies were observed in the analyses discussed in this section: (1) plural forms of a word are on average more difficult to recognize correctly than singular forms of the same word, (2) listeners were prone to omitting the final /s/ or /n/ in the plural forms but rarely did they insert a final /s/ or /n/ in the singular forms of the words, and (3) the singular-plural differences are more common for the bisyllabic words than for the monosyllabic words. It is unknown if the incorrect recognition of the plural words was more related to audibility of the final consonant or to listeners that were inattentive to small phonetic cues that did not affect word meaning. The substantial reduction in final phoneme deletion errors for the words that end in /n/ or /s/ but that do not retain meaning when omission of the final phoneme occur support the latter. However, all of the bisyllabic plural words had stress in the penultimate syllable (grave words), which inherently reduces the amount of stress exerted on the final phoneme. The fact that the singular-plural recognition differences were less pronounced in the monosyllables than in the bisyllables could serve as an argument in support of the audibility theory since in monosyllables the final phoneme is within the stressed syllable. The finding of poorer recognition performances on words that retain their meaning upon deletion of the final phoneme may offer insight into the unexplained results from two studies that observed that the Berruecos and Rodriguez word lists<sup>8</sup> had lower performances

than similar word lists recorded by the same speakers.<sup>31,32</sup> The words from the Berruecos and Rodriguez list are familiar to most Spanish dialects; however, 26% of them are forms of the word or conjugations that end in /n/ or /s/ and retain meaning and stress pattern if the final phoneme were omitted. Other comparable Spanish word lists have substantially fewer of these word forms. If 26% of the Berruecos and Rodriguez words had reduced performances compared with the other words in the lists, then overall recognition performance on the word list would be reduced, which is what both the Kroes and the Flores and Aoyama studies observed. An awareness of the reduced intelligibility of these classes of words underscores the importance of knowing the recognition characteristics of each word before a word list is compiled from the corpus.

### Penultimate Syllable and Last-Syllable Stress

Another factor that demonstrates an effect on bisyllabic-word intelligibility is syllabic stress.<sup>33</sup> Of the current 359 bisyllabic words, 318 were stressed on the penultimate syllable (the first syllable in bisyllabic words) and 41 were stressed on the last syllable. The recognition performance functions for these two groupings of bisyllabic words are shown in ►Fig. 3 (lower right panel). The polynomial 50% points and the slopes at the 50% points were 2.8-dB HL (6.6%/dB) and 7.5-dB HL (5.1%/dB) for stress on the last and penultimate syllables, respectively. This observation that words stressed on the penultimate syllable are more difficult to recognize than words stressed on the last syllable is congruent with the findings from Black<sup>33</sup> for English bisyllables and is acknowledged as a contributing variable to the overall variability associated with Spanish word-recognition performance.

### Female- and Male-Gender Words

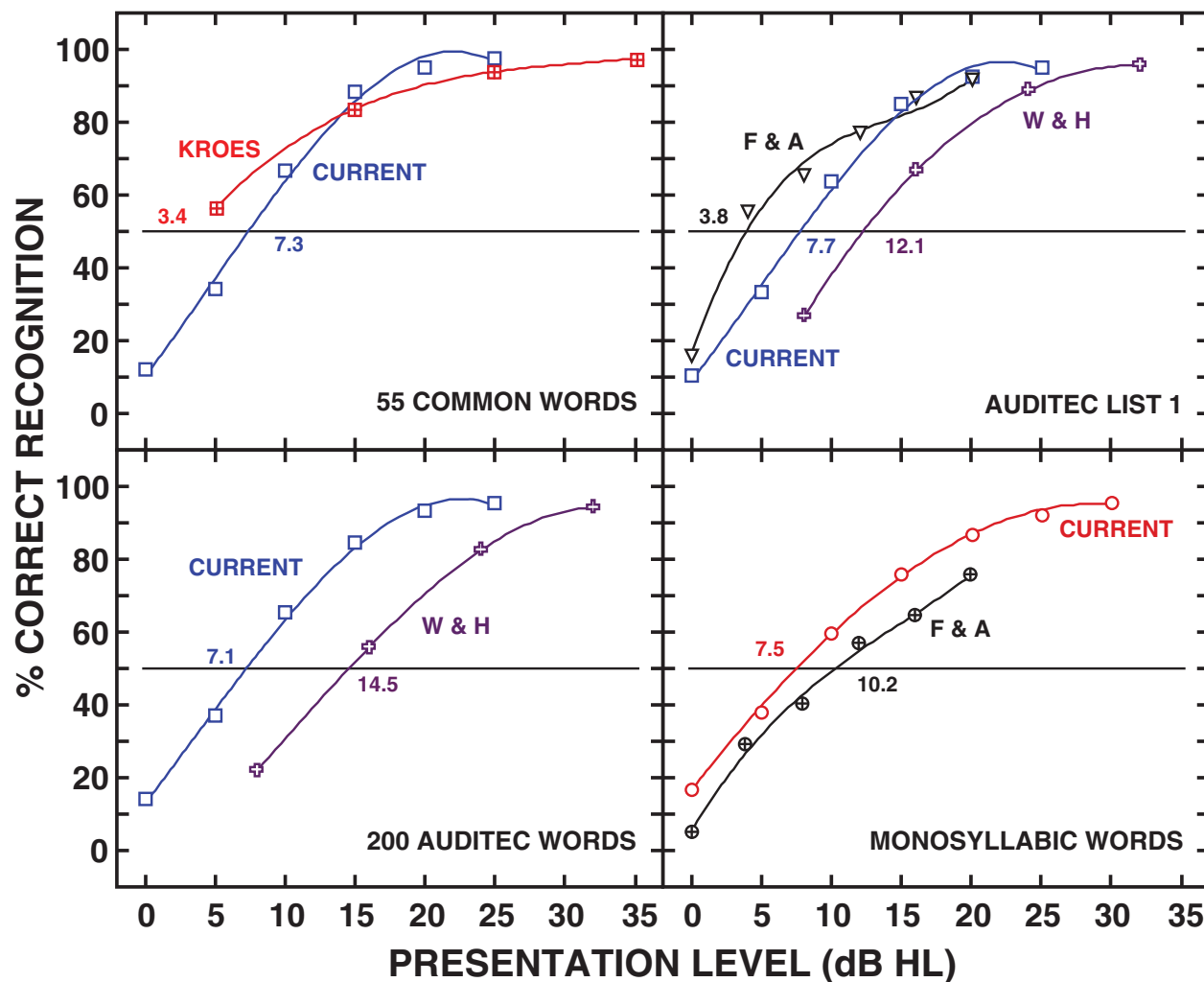
In Spanish, most nouns, adjectives, and articles are gender specific. Gender in nouns and adjectives is indicated by the final (vowel) phoneme or by the article that precedes the word. The final phoneme /a/ is commonly, but not exclusively, used to indicate the female-gender form of the word whereas a final /o/ indicates the male form. For example, *hijo* translates to *son*, whereas *hija* translates to *daughter*. Spanish verbs are not gender-specific. However, present tense verbs in singular form often end in /o/ or /a/. Specifically, when conjugated in the first person, with few exceptions, they end in /o/. When conjugated to the third person, the final phoneme of the verb most often changes from /o/ to /a/. For example, for the verb *jugar* (play), *I play* translates as *yo juego*, whereas *he plays* translates to *el juega*. Thirteen pairs of bisyllabic words (five verbs, five adjectives, and three nouns) that differed only by a change of final phoneme from /a/ to /o/ were selected for analysis. Although, as mentioned earlier, verbs are not gender-specific, for the purposes of this analysis the terms female- and male-gender words was utilized to describe words ending in /a/ and /o/, respectively.

The recognition functions for the 13 bisyllabic gender word pairs are shown in Figure S11, which also includes the mean gender word-pair data. Of the 13 words with ending vowels of /a/ (female) and /o/ (male), only one word pair, *besa*

and *beso*, failed to reach recognition performances above 60 to 70%. The mean gender word-pair functions ( $n = 13$ ) differed by 2.6 dB at the 50% point with the better performances on the male-gender words (7.0-dB HL) than on the female-gender words (9.6-dB HL). The slopes for the mean functions at the 50% points were 5.4 and 5.1%/dB, respectively, for the female- and male-gender words. Ten of the 13 word pairs had better performances at the 50% points on the male-gender words than on the female-gender words by  $\geq 3.6$  dB, with four of those word pairs having differences  $> 5$  dB. The three word pairs with better performances on the female-gender words (*barca, barco; hija, hijo; and roja, rojo*) were only better than the performances on the male counterparts by  $< 1$  dB. The conclusion is that recognition performances on most of the gender word pairs were essentially the same but on four of the word pairs (*besa, beso; juega, juego; mala, malo; and vota, voto*) notably better recognition performances were obtained on the male gender than on the female gender, ranging from 5.1 to 8.7 dB. The interesting question is whether these differences were owing to the female, male gender ending of the word, familiarity differences, the particular utterance of each word, or a combination of factors.

### Comparison with Previous Data

The final interest of this stage of the project was to compare the psychometric properties of the current recorded materials to the same materials recorded and studied by previous investigators. Such comparisons are not definitive, especially in view of the study variables that are unknown or unquantifiable. Probably the most important of these variables are inherent perceptual speaker differences, which are next to impossible to quantify, and the presentation level that on the surface is straightforward but in reality is elusive, especially when consideration is given serially to (1) how the word amplitudes were determined and set during and following the recording session (e.g., vu meter, and rms), (2) how the presentation level was quantified (e.g., sensation level or HL), and (3) what reference sound-pressure level for speech was used. With an awareness of these limitations, recognition-performance functions from several earlier studies are presented in ►Fig. 6 with the numbers in each panel noting the HL (dB) of the 50% point, which was calculated from the polynomial equation used to describe the data (note: the data for these studies are listed in Table S26). The upper left panel shows the recognition functions for 55 bisyllabic words from the current study and for the same 55 words recorded and studied by Kroes.<sup>31</sup> The higher presentation levels of the two sets of words produced almost identical recognition performances. At the 50% points the functions differ by 4 dB with the Kroes data exhibiting better performance. Unfortunately, the Kroes data did not extend below 50%, prohibiting comparisons of data at the lower presentation levels and of the slopes of the functions. The upper right panel in ►Fig. 6 contains data from the 50 words in the Auditec List 1 spoken by the Auditec speaker,<sup>10</sup> by the Flores and Aoyama speaker,<sup>32</sup> and by the speaker in the current study. The adjacent 50% points differed by approximately



**Fig. 6** Comparisons of the recognition-performance functions for various combinations of bisyllabic words from the current study (1) with the same 55 words from Kroes,<sup>31</sup> (2) with the 50 words from Auditec List 1 from Weisleder and Hodgson (W & H<sup>10</sup>) and Flores and Aoyama (F & A<sup>32</sup>), (3) with the 200 Auditec words from Weisleder and Hodgson, and (4) with the 50 monosyllabic words from Flores and Aoyama (the current study sample had 32 common words). The numbers in each panel are the hearing levels (dB) at which the 50% points occurred, calculated from the polynomial equations for the respective functions.

4 dB, with the performance on the Flores and Aoyama version 3.9 dB better than the performance on the current study, which in turn was 4.4 dB better than the performance on the Weisleder and Hodgson version. The slopes of the recognition functions at the 50% point were similar, ranging from 6.4 to 5.0%/dB. The functions in the lower left panel of ►Fig. 6 are for the 200 words of the four Auditec lists from the Weisleder and Hodgson study and from the current study; comparisons of the four 50-word lists are depicted in Figure S15. At the 50% point, the current version (7.1 dB) was 7.4 dB easier than the Weisleder and Hodgson version (14.5 dB); the slopes of the functions were similar with the slope of the current being a little steeper, 5.0 versus 4.1%/dB. As indicated in the introduction of this paragraph, these differences probably are attributable to a variety of unknown and known variables. Finally, from ►Fig. 6, recognition performances on monosyllabic words from the Flores and Aoyama study and from the current study are depicted in the lower right panel. For this comparison, there were 32 common words in

the two studies but because individual word performances were not given in the Flores and Aoyama study, the performance comparison is made between their 50 words and the 32 common words in the current study. The observation from the data are that the functions are quite similar with the 50% points separated by 2.7 dB; the slopes of the functions differed by 1.1%/dB with the function for the current function steeper, 4.0 versus 2.9%/dB. The overriding conclusion from the data in ►Fig. 6 is that considering the multitude of issues involved in comparing recognition performances among studies with different speakers and different techniques, the current recognition data are certainly “in the ballpark” with data from several previous studies.

## Conclusions

The data from the current study substantiated that there were significant differences among the recognition-performance functions for Spanish monosyllabic, bisyllabic, and trisyllabic

words both in terms of the hearing level (dB) at which the 50% points were established and of the slopes of the functions at the 50% points. Although the mean data were significantly different, there was substantial overlap among the recognition functions of the individual words from the three syllabic categories with 89% of the monosyllables and 79% of the trisyllables having 50% points within  $\pm 2$  SDs of the bisyllabic word mean 50% points. Thus the monosyllabic and trisyllabic words were as easy as or easier to understand than were many of the bisyllabic words; conversely, many of the monosyllabic and trisyllabic words were as difficult as or more difficult to understand than were many of the bisyllabic words.

The goal of this study was to establish the psychometric characteristics of Spanish monosyllabic, bisyllabic, and trisyllabic words that could serve as selection guidelines for subsequent efforts (with individuals with sensorineural hearing loss) to establish the variety of word-recognition materials for use in audiologic evaluations. The current data will be used to govern the inclusion of words especially with respect to audibility and the influences that other characteristics like familiarity, form, syllabic stress, and gender have on word-recognition performance. The reader is reminded that the characteristics reported here only apply to the particular utterances of the materials used in this study of young adult listeners with normal hearing sensitivity.

## Acknowledgments

The authors acknowledge the contributions made by Dr. Lillian Pintado in the audio-editing portion of this project.

### Notes

Parts of this manuscript were presented at the Puerto Rico Academy of Audiology Convention. Río Grande, Puerto Rico (February, 2016), and the American Academy of Audiology Conventions in Phoenix, Arizona (April, 2016) and in Columbus, Ohio (March, 2019).

### Funding

This work was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Grant R25MD007607. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### Conflicts of Interest

Dr. Carlo reports a grant from the NIH, during the conduct of the study. Dr. Wilson reports a grant and salary from his employer, the U.S. Department of Veterans Affairs, during the conduct of the study. Dr. Villanueva-Reyes reports no conflict of interest.

## References

- Hudgins CV, Hawkins JE, Karlin JE, Stevens SS. The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope* 1947;57(01):57–89
- Egan JP. Articulation testing methods. *Laryngoscope* 1948;58(09): 955–991
- Tato JM, Lorente F, Belo JA, Tato JM. Características acústicas de nuestro idioma. *Rev Otorrinolaringol* 1948;1:17–34
- Cancel-Ferrer C. Hearing test (social adequacy index) for Spanish-speaking people (unpublished master's thesis). Ann Arbor, MI: University of Michigan; 1951
- Ferrer O. Speech audiometry: a discrimination test for Spanish language. *Laryngoscope* 1960;70(11):1541–1551
- Rosenblüt B, De Cruz J. Listas de palabras en español para pruebas de discriminación. *Rev Otorrinolaringol* 1962;22(02): 37–49
- Quirós J. Material de listas de sílabos y palabras para logoaudiometría. Buenos Aires, AR: La Escuela Superior de Fonoaudiología Corrientes; 1965:1723
- Berrucos TP, Rodríguez JL. Determination of the phonetic percent in the Spanish language spoken in Mexico City, and the formation of PB lists of trochaic words. *International Audiology* 1967;6(02): 211–216
- Zubick HH, Irizarry LM, Rosen L, Feudo P Jr, Kelly JH, Strome M. Development of speech-audiometric materials for native Spanish-speaking adults. *Audiology* 1983;22(01):88–102
- Weisleder P, Hodgson WR. Evaluation of four Spanish word-recognition-ability lists. *Ear Hear* 1989;10(06):387–392
- Whitley MS. Spanish/English Contrasts: A Course in Spanish Linguistics. 2nd ed. Washington, DC: Georgetown University Press; 2002
- Algeo J. What consonant clusters are possible? *Word* 1978;29(03): 206–224
- Bedore L. The acquisition of Spanish. In: Taylor O, Leonard L, eds. *Language Acquisition Across North America: Cross-Cultural and Cross-Linguistic Perspectives*. San Diego, CA: Singular; 1999:157–207
- Guirao M, García-Jurado M. Frequency of occurrence of phonemes in American Spanish. *Revue Québécoise de Linguistique* 1990;19 (02):135–149
- Jullian A, Chang-Rodríguez E, eds. *Frequency Dictionary of Spanish Words*. Hague, The Netherlands: Mouton; 1964
- Fucks W. Die mathematischen Gesetze der Bildung von Sprachelementen aus ihren Bestandteilen (The mathematical laws of the formation of language elements from their components). *Nachrichtentechnische Fachberichte* 1956;3:7–21
- McCullough JA, Wilson RH, Birck JD, Anderson LG. A multimedia approach for estimating speech recognition of multilingual clients. *Am J Audiol* 1994;3(01):19–22
- Bishop A. Psychometrically equivalent bisyllabic word-lists for word recognition testing in Spanish (unpublished master's thesis). Provo, UT: Brigham Young University; 2009
- American Standards Association. *Volume Measurements of Electrical Speech and Program Waves, C16.5–1954*. New York, NY: American Standards Association; 1954
- Killion MC. Comparison of vu-meter-based and rms-based calibration of speech levels. *J Acoust Soc Am* 2009;126(03):EL97–EL99
- Adobe Systems, Inc. *Adobe Audition CCC*. San Jose, CA; 2015
- Chinn HA, Gannett DK, Morris RM. A new standard volume indicator and reference level. *Bell Labs Tech J* 1940;19(01):94–137
- Wilson RH. Amplitude (vu and rms) and temporal (ms) measures of two Northwestern University Auditory Test No. 6 recordings. *J Am Acad Audiol* 2015;26(04):346–354
- American National Standards Institute. *Specification for Audiometers ANSI S3.6–2010*. New York, NY: American National Standards Institute; 2010
- Wilson RH, Sanchez VA. Effects of the carrier phrase on word recognition performances by younger and older listeners using two stimulus paradigms. *J Am Acad Audiol* 2020;31 (06):412–441
- Wilson RH, Margolis RH. Measurement of the auditory thresholds for speech stimuli. In: Konkle D, Rintelmann W, eds. *Principles of*

- Speech Audiometry. Baltimore, MD: University Park Press; 1983: 79–126
- 27 Finney DJ. Statistical Method in Biological Essay. London: C. Griffen; 1952
- 28 Dirks DD, Takayanagi S, Moshfegh A, Noffsinger PD, Fausti SA. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear Hear* 2001;22(01):1–13
- 29 Shi L-F, Sánchez D. The role of word familiarity in Spanish/English bilingual word recognition. *Int J Audiol* 2011;50(02): 66–76
- 30 Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear Hear* 1998;19(01):1–36
- 31 Kroes PA. Comparison of the relative intelligibility of four Spanish speech-audiometric tests (unpublished doctoral dissertation). Columbus, OH: Ohio State University; 1973
- 32 Flores L, Aoyama K. A comparison of psychometric performance on four modified Spanish word recognition tests. *Texas Journal of Audiology and Speech-Language Pathology* 2008;31:64–70
- 33 Black JW. Accompaniments of word intelligibility. *J Speech Hear Disord* 1952;17(04):409–418