

Reynita R. Sagon, BS, AuD (student)¹,
Rosalie M. Uchanski, PhD^{1,2}

Washington University in St. Louis - School of Medicine

¹Program in Audiology and Communication Sciences

²Department of Otolaryngology

Saint Louis, Missouri USA

The Development of Ilocano Word Lists for Speech Audiometry

ABSTRACT

Objective: The goal of this work is the creation of word lists, in Ilocano, suitable for use in speech audiometry.

Methods: First, estimates of the distribution of speech sounds and of the most common syllable structures in Ilocano were found from a phonetic transcription analysis of nearly 3000 words obtained from three magazine articles. Second, 372 two-syllable words were rated, for commonness, by fifteen native speakers of Ilocano who currently reside in Hawai'i. Finally, various combinations of two-syllable words were made to produce 50-item lists.

Results: First, an estimate of the distribution of speech sounds in Ilocano was found, with frequencies of occurrence ranging from 22.4%, for the speech sound /a/, to 0.007%, for the speech sound /v/. The syllable-structure analyses revealed that a very small number of distinct monosyllabic words were used very frequently. Two-syllable words were also used frequently, but were attributed to many distinct words. Second, from the rating results, approximately 70% of the two-syllable words in the rating survey were judged as common by 12 or more of the raters.

Finally, four lists of 50 words each were constructed using only common two-syllable words with the most frequent two-syllable structures found in Ilocano. Each word list has a distribution of speech sounds that approximates that found from the phonetic analysis, and hence each list is roughly phonetically-balanced.

Conclusions: These word lists may be of value to otolaryngologists and audiologists who work with native speakers of Ilocano.

Keywords: *Ilocano, Ilokano, phonetically-balanced, speech audiometry, word lists*

A COMPREHENSIVE audiological examination consists of pure-tone audiometry, speech audiometry, and physiological assessments deemed necessary by the results of behavioral testing. Pure-tone audiometry employs specific tonal stimuli to determine the severity and configuration of a patient's hearing loss while speech audiometry consists of word stimuli to determine a patient's speech reception threshold (SRT) and word recognition score/speech discrimination score (WRS). Speech audiometry may also provide the clinician with valuable diagnostic evidence for the evaluation and management of hearing disorders^{1,2}. For example, word recognition tests may be used to assess retro-cochlear function by testing whether intelligibility scores decrease significantly as intensity increases³. Speech audiometry is often used to provide an indication of the functionality of a person's entire speech perception system and, therefore, assess a patient's communication ability⁴. The accuracy of this testing, however, depends on the patient's knowledge of the speech material. Testing a patient in his/her non-native language may yield inaccurately low scores because foreign words may appear as nonsense stimuli^{5,6}. To take full advantage of the assessment and diagnostic uses of speech

Correspondence: Rosalie M. Uchanski, PhD
Washington University in St. Louis – School of Medicine
Department of Otolaryngology
Campus Box 8115
660 South Euclid Avenue
Saint Louis, MO USA 63110
314-747-7191 (voice)
314-747-7230 (fax)
Email: uchanski@ent.wustl.edu
Reprints will be available from the author.

No funding support was received for this study. The authors signed a disclosure that they have no proprietary or financial interest with any organization that may have a direct interest in the subject matter of this manuscript, or in any product used or cited in this study.

audiometry, the patient should be administered speech material in his/her native language⁷⁻¹¹. Care must also be taken to insure that the speech material is not too familiar. For example, when a patient obtains an annual comprehensive audiometric examination inflated scores may be obtained due to repeated exposure with the speech material (i.e., use of a single, repeated word list). To prevent over-familiarization, such speech material should have multiple word lists available so that lists can be alternated in an attempt to prevent memorization of any specific words. Ideally, word lists for speech audiometry should consist of words that are familiar, though not overly-familiar to the patient, and these lists should consist of words from the patient's native language or dialect.

There are an estimated 8 million speakers of Ilocano worldwide, though most reside in the Philippines particularly in Northern Luzon. In the Philippines, Ilocano is the third most common language after Tagalog and Cebuano, and is spoken by roughly 10% of the population. Significant populations of Ilocano speakers can also be found in various countries and regions including Singapore, Hong Kong, Kuwait, Canada, Europe and the United States. In the United States, a high concentration of Ilocano speakers can be found in Hawai'i, California, Washington and Alaska¹²⁻¹⁴.

Currently, however, there are no known speech audiometry tests in the Ilocano language. To address the need expressed by Carhart¹⁵ of developing speech materials in all spoken languages, this study is focused on the development of Ilocano word lists for speech audiometry.

General issues in the development of word lists for speech audiometry

Several classic articles describe some main principles for speech audiometry^{6,15-17}. While many factors have been identified that can affect the results of speech tests, the focus of this discussion will be on those factors related to the choice of speech material. Hence, factors such as presentation level, type of presentation (live-voice vs. recorded-voice) and choice of talker^{16,18,19} will not be addressed further. Additionally, there is general agreement that words for assessing a speech reception threshold (SRT) and words for use in word recognition tests (to obtain a WRS) should be selected using different criteria^{4,20}. Since the primary goal of this study is the development of word lists for word recognition tests, only word-selection criteria relevant for measuring a WRS will be discussed.

For every spoken language, three main factors for selecting words for use in a word recognition test seem to be i) good representation of the phonetic structure, ii) good representation of the word-level syllable structure, and iii) word commonness or familiarity. The first factor, good representation of the phonetic structure of the language, is generally addressed by the creation of word lists that are phonetically balanced (e.g., CID W-22 lists¹⁷; CNC lists⁵). A phonetically-balanced (PB) word list should have a proportional representation in each list of the sounds that occur in everyday speech for that language²¹. Presumably, the use of phonetically-balanced word lists contributes to their validity for assessing a person's ability to understand ordinary conversation.

Table 1: Consonant sounds in Ilocano. The consonants are organized by their place and manner of articulation (rows) and by their voiced vs. unvoiced method of production (columns).

CONSONANTS		
	voiced	unvoiced
BILABIAL		
plosives	b	p
	b̥	p̥
nasals	m	
	m̥	
LABIODENTAL		
	v	f
DENTAL		
plosives	d	t
	d̥	t̥
nasals	n	
	n̥	
trill	r	
fricatives		s
		s̥
affricates	ʈ	ʈ̥
postalveolar		ʃ
lateral approximate	l	
	l̥	
PALATAL		
approximate	j	
	j̥	
VELAR		
plosives	g	k
	g̥	k̥
nasals	ŋ	
	ŋ̥	
LABIO-VELAR		
approximate	w	
GLOTTAL		
plosives	ʔ	
fricatives		h

MONOPHTHONGS				
	FRONT	CENTRAL	BACK ROUND	BACK UNROUND
HIGH	i			u
MID	e			o
LOW		a		
DIPHTHONGS				
	ay			
	iw			
	uy			
	oy			
	aw			

Table 2. Vowel sounds in Ilocano. Monophthongs and diphthongs are listed separately.

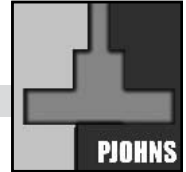


Table 3. Frequency of occurrence (expressed in percent) of each speech sound in Ilocano for this language sample (total number of speech sounds: 14,327). The sounds are listed in descending order, by percent occurrence, for consonant and vowel sounds separately. Please note: For this language sample, the frequency of occurrence is exactly zero for both /j:/ and /f:/.

consonant sound	Percent (%) occurrence	vowel sound	Percent (%) occurrence
t	8.341	a	22.356
n	7.238	i	12.417
k	4.886	u	3.671
d	3.623	o	3.413
s	3.616	e	3.413
g	3.525	ay	1.933
p	3.176	aw	0.565
m	2.750	oy	0.328
l	2.555	uy	0.077
ŋ	2.464	iw	0.035
r	2.359	ALL VOWELS	48.210
ʔ	2.087		
b	1.794		
ɳ	0.796		
j	0.635		
w	0.614		
ɖ	0.307		
h	0.265		
ʃ	0.168		
m:	0.168		
ɕ	0.112		
k:	0.077		
ɹ	0.077		
ŋ:	0.028		
ɹ:	0.028		
ɕ:	0.028		
ɹ̄	0.021		
h:	0.021		
ɕ:	0.014		
ɹ:	0.014		
v	0.007		
ɹ:	0.000		
f	0.000		
ALL	51.790		
CONSONANTS			

And, use of phonetically-balanced word lists ensures that the speech sounds of a given language will be presented acoustically to native listeners during a speech test. It is not clear, though, exactly how to determine the distribution of sounds that occur in the everyday speech of a language. Distributions can depend on a talker's dialect, on the source of the spoken material (e.g., text that is read vs. spontaneous conversation) and on the topic of conversation^{16,22,23}. For example, in American English there is a much greater occurrence of the diphthong sound /a/ (pronunciation of the pronoun "I") in spontaneous conversation than in spoken narrative or 'read' text²³. The necessity of using word lists that are phonetically-balanced has been questioned recently by Martin *et al.*²⁴. They found, for WRS's, that "the total score based on randomly selected words is not substantially different from the total score based on carefully selected, PB word lists" (p. 492). While these results challenge the necessity of designing and using phonetically-balanced word lists, their results do not imply that the use of phonetically-balanced word lists provides an inaccurate or incorrect assessment of speech communication ability. And, Martin *et al.* did not examine the phonetic balance of their randomly selected words; it is possible that their randomly-selected word lists have phonetic distributions similar to those of the phonetically-balanced lists.

The second factor, good representation of the word-level syllable structures, has received somewhat less attention than has phonetic balance. However, as argued for phonetic balance, the use of word lists with a good representation of the syllable structure of a language presumably addresses and satisfies the validity of such word lists for audiometric speech tests. In English, most word lists for word recognition tests are monosyllabic, as the most common words in English are monosyllabic²³. Other languages have different types of syllable structures for common words and different frequencies of occurrence of each syllable type. Consequently, for many languages, e.g., Spanish, Brazilian Portuguese, Italian and Mandarin, word lists developed for speech audiometry consist of polysyllabic words^{9-11,20,25,26}.

The third factor to consider for the design of words lists is word commonness or familiarity¹⁷. As mentioned previously, word lists for speech audiometry should consist of words that are familiar though not overly-familiar to the patient. Presumably, through the use of words that are generally familiar to all adults, any effects due to differences in educational background should be minimized. That is, words selected for a word list should be familiar to as many adults as possible regardless of their level of formal education. The effect of word familiarity on speech intelligibility has been studied for American English²⁷⁻³¹, and this effect can be substantial especially if practice is not provided³². Rare or infrequently-used words are identified with poorer accuracy than are common or frequently-used words^{29,31}. Though most of these results are from adults with normal hearing listening to words degraded by noise or by low-pass filtering, adults with hearing loss should exhibit similar effects due to word familiarity or commonness. More recently, the concept of 'lexical difficulty' was introduced³³⁻³⁵ for the development of the 'easy' and 'hard' Lexical Neighborhood Tests. Commonness (frequency of occurrence) of a word and the density of a word's 'neighborhood' both contribute to a word's 'lexical difficulty.' 'Easy' words are common and have a relatively sparse neighborhood

(i.e., there are few words with speech sound sequences similar to that of the target ‘easy’ word) while ‘hard’ words are less common and have relatively dense neighborhoods (i.e., there are many words with a similar speech sound sequence, differing from the target ‘hard’ word by only one speech sound). ‘Lexical difficulty’ has also been found to have a large effect on WRS’s³⁵. If the test population is young children, not adults, then additional restrictions should be considered regarding commonness or familiarity. In particular, the vocabulary of a word list should not exceed that expected of a child listener, and a child’s expected vocabulary depends substantially on chronological age and experience with spoken language. Such additional considerations for speech audiometry for a child population will not be discussed further here.

There are certainly other criteria that could be used in the design of speech materials or word lists for speech audiometry. For example, the Danish speech material, *Dantale*, explicitly controls the parts of speech (nouns, adjectives and verbs) in their lists, and intentionally avoids emotional or objectionable words³⁶. However, we have chosen to focus on the three main factors described above as these seem to have greatest validity.

In summary, our goal is to develop multiple word lists in Ilocano for possible use in speech audiometry. The word lists should be phonetically-balanced (i.e., reflect the distribution of speech sounds in Ilocano), reflect the distribution of word-level syllable structures in Ilocano, and consist of common or familiar words.

There are three main parts to this study, namely 1) an estimate of the distribution of speech sounds and word-level syllable structures in Ilocano, 2) ratings of commonness (word familiarity), and 3) creation of multiple word lists. Since the method for the third part (creation of word lists) depends on results from the first two parts, three method-plus-results sections will be presented sequentially.

Part 1: Estimate of the distribution of speech sounds and word-level syllable structures

METHOD

To create word lists that are phonetically balanced and that reflect the syllable structures in Ilocano, the distribution of speech sounds and syllable structures must be known or estimated.

While there are a handful of Ilocano dictionaries in print¹⁴, there are no known reports or articles on the frequency of occurrence of speech sounds or the frequency of occurrence of syllable structures in Ilocano.

The phonetic distribution of the Ilocano language was estimated from three articles in *Bannawag*, a weekly magazine written, printed and distributed in the Philippines. All articles were written in the Ilocano language and discuss different topics. The three articles are entitled “Ina ti Iluko iti Unibersidad ti Hawaii,” “Ti Kaso ti Di Makita a Seradura” and “Intayon Agbutos.” “Ina ti Iluko iti Unibersidad ti Hawaii” was published on May 17, 2004, and the title translates to “the mother/founder of Ilocano at the University of Hawaii.” The article discusses a specific professor, Professor Prescila Espiritu, who helped develop an Ilocano program (including language, dance, drama and culture) and

Table 4. Number of syllables in each of the 2918 words and in each of the 889 distinct words in the language sample.

# OF SYLLABLES (in the word)	# OF WORDS (with that # of syllables)	# OF DISTINCT WORDS (with that # of syllables)
1	977	30
2	811	188
3	539	265
4	366	214
5	146	120
6	67	60
7	10	10
8	2	2
TOTAL	2918	889

Table 5. Frequency of occurrence (expressed in percent) for each syllable structure for the distinct two-syllable words in this language sample. The six syllable structures that account for most (~83%) of the two-syllable words are indicated with bold font.

Syllable Structure	# OF DISTINCT WORDS (with that syllable structure)	Percent (%) occurrence
CCVCCV	5	2.66
CCVCCVCC	1	0.53
CCVCV	5	2.66
CCVCVC	2	1.06
CVCCCV	1	0.53
CVCCV	25	13.3
CVCCVC	15	7.98
CVCV	34	18.1
CVCVC	46	24.5
CVV	2	1.06
CVVC	6	3.19
VCCCV	2	1.06
VCCVC	3	1.60
VCCVC	5	2.66
VCV	19	10.1
VVC	16	8.51
VV	1	0.53
TOTAL	188	100.0

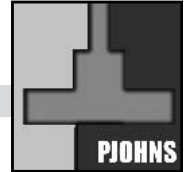


Table 6. Number of words rated as “common,” by various numbers of raters. A total of 372 words were rated by each rater, but 3 words (“bukno,” “dulin,” and “essem”) were duplicated in the two survey sub-lists. Thus, a total of 369 distinct words were rated.

EXACT # OF RATES (who judged word as ‘common’)	# OF WORDS (with that exact # of raters who judged word as ‘common’)
15	73
14	74
13	71
12	40
11	24
10	21
9	14
8	15
7	16
6	8
5	6
4	4
3	0
2	0
1	2
0	1
TOTAL	369

currently teaches Ilocano at the University of Hawai‘i. “Intayon Agbutos” or “Let’s vote,” was published on April 26, 2004. This article encourages Philippine citizens to vote, and discusses issues that may hinder citizens from voting.

The last article, “Ti Kaso ti Di Makita a Seradura” or “The Case of the Hidden Door” was published on June 7, 2004. This article discusses religion. As mentioned in the introduction, any estimate of the distribution of speech sounds will depend on the source of the language sample.

Ideally, the language sample should be spoken, should include a wide variety of spoken samples (such as radio announcements, television broadcasts, religious sermons, and conversations on various topics with talkers of multiple age groups), and should be extremely large. Though this language sample is based on written, not spoken, narrative, it does encompass several different topics, and it is of moderate size.

Each sentence in each article was written on a separate sheet and transcribed phonetically by the first author (a native speaker of Ilocano) using a dictionary pronunciation^{14,37}. A list of speech sounds in Ilocano was obtained from Ilocano Dictionary and Grammar Ilocano-English, English-Ilocano³⁷. Twelve speech sounds, /p:/, /b:/, /m:/, /n:/, /s:/, /l:/, /j:/, /k:/, /g:/, /ɲ:/, /ɳ:/, /ɰ:/, /kɰ:/, /gɰ:/ and /ŋ:/, were added to this list after the transcription of the language sample was completed. These speech sounds differ from /p/, /b/, /m/, /n/, /d/, /n/, /s/, /l/, /j/, /k/, /g/ and /ŋ/, respectively, in duration. The phonetic symbol (:̣) indicates a longer duration, and this lengthening is called ‘geminate’ by linguists. These twelve geminate

consonants are considered distinct consonants, rather than allophonic variants, because the lengthening of these speech sounds changes the literal meaning of a word. For example, the word “paddak,” transcribed /paddak/ means “to step” while “padak,” transcribed /pad:ak/ means “similar to me.”

Transcribed words were then listed in alphabetical order and tallied. Tabulations were made of the number of occurrences of each speech sound in each word, and tabulations were also made of weighted occurrences of each speech sound. The weighted occurrence accounts for the number of times a specific word occurred in the language sample. In sum, our estimate of the distribution of speech sounds in Ilocano is based on the number of occurrences of each speech sound in each word and on the number of occurrences of each word in this language sample of three magazine articles.

In an analogous manner, an estimate of the distribution of word-level syllable structures was found for this language sample. Word-level syllable structures were tabulated, again by the first author, using a dictionary pronunciation of each word, and then by transcribing the pattern of consonants and vowels in each word. For example, the words “essem” and “bukno” were assigned the word-level syllable structures “VCVC” and “CVCCV,” respectively. In a manner similar to that for speech sounds, the distribution of word-level syllable structures across the language sample was estimated.

RESULTS

The three magazine articles used for our transcriptions contained 2,918 words of which 889 were distinct words. In units of speech sounds, this same sample contained 14,327 speech sounds, of which 43 were distinct. Specifically, the Ilocano language consists of the following speech sounds: /p/, /p:/, /b/, /b:/, /m/, /m:/, /n/, /n:/, /s/, /s:/, /l/, /l:/, /j/, /j:/, /k/, /k:/, /g/, /g:/, /ɲ/, /ɲ:/, /ɳ/, /ɳ:/, /ɰ/, /ɰ:/, /kɰ/, /kɰ:/, /gɰ/, /gɰ:/, /ŋ/, /ŋ:/, /w/, /w:/, /ʔ/, /ʔ:/, /h/, /h:/, /ɰ̣/, /ɰ̣:/, /ɳ̣/, /ɳ̣:/, /i:/, /i:/, /u/, /u:/, /a/, /a:/, /ay/, /ay/, /oy/, /oy/, /uy/, and /aw/: 33 consonant sounds (Table 1) and 10 vowel sounds (Table 2) of which 5 are monophthongs and 5 diphthongs. The consonants /f/ and /v/ (listed in Tables 1 and 3) were, however, eliminated from further analysis because these sounds are usually only found in proper names or words borrowed from the Spanish language. For subsequent analyses and discussion, we will consider only 41 distinct speech sounds (*i.e.*, after the elimination of /f/ and /v/). Based on this language sample, the frequency of occurrence of each speech sound was calculated and these values (expressed as percents) are listed in Table 3, in descending order. Compared to spoken English, the split between consonant and vowel sounds is somewhat different. Denes²² reports that approximately 60.7% of speech sounds are consonants and 39.3% are vowels for English. Berruecos and Rodriguez²⁸ report roughly 56% of speech sounds are consonants and 44% are vowels, for Spanish spoken in Mexico City. For Ilocano, we find a more equal split between consonant (51.8%) and vowel (48.2%) sounds. Within the consonant group, our results are similar to those of Denes for English and of Turrini *et al.*⁹ for Italian, in that /t/ and /n/ are the two most common consonant sounds, and /ɰ̣/ and /kɰ̣/ are two of the least common consonant sounds (ignoring the Ilocano geminate consonants).

To examine syllable-structure, our initial step was to count the number of syllables in all the words in this language sample. In the second column of Table 4, the data show that about one third (977 of 2918) of the words in the language sample are monosyllabic. However, these 977 occurrences of one-syllable words are due to only 30 distinct words. That is, very few different monosyllabic words occurred, though each word occurred numerous times. As a consequence of the small number of distinct monosyllabic words in the language sample, monosyllabic words were not considered further for the development of word lists in Ilocano. Similar to word lists for other languages, two-syllable words will be used^{9,10,20}. The second step was to examine the syllable structure, that is, the sequence of consonants and vowels, of the 188 distinct, or unique, two-syllable words. The results of this examination are shown in Table 5. Seventeen syllable structures were found for these two-syllable words. Six syllable structures account for most of the two-syllable words (82.5%), namely: CVCCV, CVCCVC, CVCV, CVCVC, VCV and VCVC.

In summary, the first part of the study yielded a full description of all the speech sounds in Ilocano, an estimate of the frequency of occurrence of each speech sound in a moderately-sized Ilocano language sample, an estimate of the frequency of occurrence of words with different numbers of syllables, and an estimate of the frequency of occurrence of various syllable structures for two-syllable words. Based on these estimates and results, two-syllable words with six particular syllable structures were chosen for developing word lists in Ilocano for speech audiometry.

Part II: Ratings of Commonness (Word Familiarity)

METHOD

As discussed in the introduction, common words are recommended for word lists for speech audiometry. *E.g.*, Hirsh *et al.*¹⁷ state, "In general, the vocabulary consists of very common words." Since no reports could be found on the familiarity or commonness of Ilocano words, a rating of word familiarity or commonness similar to that of Hirsh *et al.*¹⁷ was conducted with native speakers of Ilocano.

A large list of Ilocano two-syllable words was created. This list consisted of all the distinct two-syllable words found in the language sample (188) and approximately 174 additional two-syllable words chosen randomly from two Ilocano dictionaries^{14,37}. This large list was divided in two sub-lists of equal size (A & B) with 186 words in each. Three words ("bukno," "dulin," and "essem") were included in both sub-lists to examine rater reliability. Additionally, to insure that raters used the full rating scale, a handful of words that the first author considered "uncommon," words such as "ebkas" and "esngaw," were also included in the survey.

Written instructions for the rating survey were provided in both English and Ilocano. Raters were asked to choose one category for each word: "common," "somewhat common" or "uncommon." One example of a "common" word and one of a "somewhat common" word were also included in the instructions. On the surveys, words were listed alphabetically by Ilocano spelling, and an English definition was

Table 7. The expected, and the actual number of occurrences of each speech sound in the four ILPB-50 word lists of Appendix B. The expected number of occurrences is based on our original estimate (see text) of approximately 230 speech sounds per word list, and on the frequency of occurrence for each speech sound found in this language sample (see Table 3).

Speech Sound	Number of occurrences					
	Expected	Actual (List 1)	Actual (List 2)	Actual (List3)	Actual (List4)	Average (4 lists)
p	7	6	5	5	8	6
b	4	13	16	14	12	14
m	6	8	6	6	6	7
t	19	17	12	12	21	16
d	8	12	9	11	11	11
ɗ:	1	0	1	0	1	1
n	17	12	16	14	6	12
ɲ:	2	1	1	0	2	1
r	5	6	10	7	6	7
s	8	13	11	12	14	13
l	6	11	9	13	18	13
j	1	0	1	2	0	1
k	11	9	9	11	11	10
g	8	14	11	14	14	13
ɰ	6	6	10	6	7	7
w	1	2	2	1	1	2
ʔ	5	3	5	3	2	3
h	1	0	0	0	0	0
i	29	18	20	22	23	21
e	8	4	6	5	4	5
u	8	18	13	11	14	14
o	8	5	5	8	4	6
a	51	47	51	44	50	48
ay	4	7	3	3	3	4
aw	1	1	2	6	2	3
oy	1	0	0	1	0	0
Total	226	233	234	231	240	235

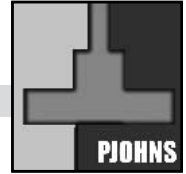


Table 8. The expected, and the actual number of occurrences of each syllable structure for the words in the four ILPB-50 lists of Appendix B. The expected number of occurrences is based on a total of 50 words in each list, and on the six most frequently occurring syllable structures (from Table 5) scaled up to make 100%.

Syllable Structure	Number of occurrences					
	Expected	Actual (List 1)	Actual (List 2)	Actual (List3)	Actual (List4)	Average (4 lists)
CVCVC	15	20	21	18	23	21
CVCVC	11	10	9	14	8	10
CVCCV	8	6	4	3	4	4
VCV	6	4	5	2	3	4
VCVC	5	5	4	6	4	5
CVCCVC	5	5	7	7	8	7
Total	50	50	50	50	50	50

provided. The first page of one of the rating surveys (*i.e.*, one of the sub-lists) is shown in Appendix A.

A total of 15 native Ilocano speakers rated these 372 words. All raters were born in the Philippines in the Northern Luzon region, and now reside in Central Hawai'i in communities where Ilocano is commonly spoken. The raters' ages range from 28 to 76 years, and there were 7 males and 8 females. The raters were given two surveys (*i.e.*, the two sub-lists) on separate days in an attempt to reduce fatigue and/or boredom. Responses were compiled for each word, for each rater and across all raters.

RESULTS

For each word, a count was made of the number of raters who judged the word as "common," as "somewhat common" or as "uncommon." Words were then separated according to the number of raters who rated the word as "common." Additionally, for each rater, the number of words given each rating was found. There was some variability across the raters; at the extremes, one person rated only 35% of the words as "common" while another rated 98% of the words as "common." However, most raters judged most of the words as "common": the median percentage across the 15 raters was 78% of the survey words were rated as "common."

In addition, the raters were generally self-consistent. For the words repeated in the two sub- lists, 13, 12, and 10 of the 15 raters gave exactly the same ratings for the three duplicated words, respectively. Finally, to account for differences amongst the raters, we decided that words rated as "common" by a large portion (12 or more; $\geq 80\%$), but not necessarily by all, of the raters would be considered "common."

Table 6 summarizes the number of words in the survey that were rated "common" by all 15 raters, by exactly 14 of the raters, by exactly 13 of the raters, etc. Using this 80% rule, a total of 258 words (73 + 74 + 71 + 40) from the rating surveys were deemed "common." However, 5 of these words, judged as "common" by most raters, have syllable structures that are not common (*e.g.*, VCCVC, VCCV, VVC and CVVC; see

Table 5). Consequently, 253 words remained that satisfy the following criteria: i) have two-syllables, ii) have a commonly-occurring syllable structure, and iii) are judged as "common" by a large portion ($\geq 80\%$) of our raters. Only these 253 words were considered for the creation of speech audiometry word lists.

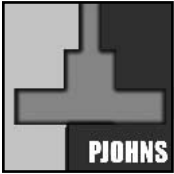
Part III: Creation of Word Lists

METHOD

The 253 two-syllable words that met the criteria listed above were placed into a 'pool' of available words. Ideally, the number of items in an audiometric word list should be large so that the margin of uncertainty for the test result is minimal⁶. And, a list should have enough items so that the full *range* of speech sounds in a language is represented¹⁵. However, a very large word list is not feasible clinically. We chose 50 as the number of items for creating Ilocano word lists. Though there is no absolute standard for the number of items in a test list, many commonly used audiometric word lists have 50 items (*e.g.*, PB-50, W-22, NU-6, CNC, LNT^{16,17,34,39,40}).

To determine which of the 41 phonemes would be included in a 50-word list the following reasoning was used. Since all words are two-syllable words, there are exactly two vowels per word, and roughly two to three consonants per word. Thus, we would expect 100 vowels (two vowels per word \times 50 words) and about 130 consonants (roughly 2-3 consonants per word \times 50), for a total of about 230 speech sounds per 50-word list. Sounds that occur with a frequency of about 0.25% will 'round up' to an expected occurrence of exactly once in a 50-item, 230-sound list ($0.44\% \times 230 = 1$ exactly, while $0.25\% \times 230 = 1$ 'rounded'). Using this frequency-of- occurrence criterion, the vowels that should be represented in the word lists are those that occurred in the language sample with a frequency greater than 0.25% (see Table 3), namely: /a/, /i/, /u/, /o/, /e/, /ay/, /aw/, and /oy/. Similarly, the consonants that should be represented in the word lists are also those with a frequency of occurrence greater than 0.25%, namely: /t/, /n/, /k/, /d/, /s/, /g/, /p/, /m/, /l/, /ʃ/, /t/, /ʔ/, /b/, /r:/, /j/, /w/, /d:/, and /h/. In summary, the 18 consonants and 8 vowels listed above should be represented in the Ilocano word lists. The 15 consonant sounds and 2 vowel sounds (see Table 3) with frequencies of occurrence less than 0.25% are explicitly excluded as they would not be expected to be represented in a 50-item word list consisting of approximately 230 speech sounds. This reasoning is analogous to that implicit in the design of the English W-22 lists in which there are no occurrences of very infrequent sounds, such as /ʒ/¹⁷.

The set of 253 words in the available "pool," *i.e.*, the set of words that met the criteria of having two syllables, being rated as "common," and having a frequently occurring syllable structure, was reduced further. Words that contained any of the infrequently occurring consonant or vowel sounds were removed from the "pool." After this reduction, only 226 words remained. From these 226 words, combinations of 50-word subsets were compiled to create lists, ideally with distributions of speech sounds that approximate that of the language sample. The first step was to alphabetize the 226 words. Then the entire list was sub-



divided into four smaller lists. List 1 was compiled by choosing the first word and every fourth word thereafter. List 2 was created by choosing the second word and every fourth word thereafter. Lists 3 and 4 were compiled using the same method. Numerous iterations involving swapping of words amongst the lists and amongst the extra words (only 200 words are used in four lists of 50) were performed manually by the first author to achieve approximate phonetic balance.

RESULTS

Four 50-item Ilocano word lists (ILPB-50) were created (see Appendix B). Table 7 is a summary of the expected number and the actual number of speech sounds in each of the four Ilocano word lists. An analysis was also made of the syllable structures present in each of these four word lists. Table 8 lists the expected and the actual number of occurrences of each syllable type for each list. In both Tables 7 and 8, there are differences between the expected and the actual number of occurrences, of speech sounds and syllable structures, respectively. One of the most noticeable differences is in the number of /b/'s in the four ILPB-50 lists (13, 16, 14, and 12) compared to the expected number (4) in each list. A similar trend is noted also for /s/, /l/, /g/ and /u/, in that the actual numbers of occurrences in the four lists are consistently greater than expected. And, an opposite trend is noted for /i/; the actual numbers of occurrences (18, 20, 22, and 23) are smaller than expected (29). This discrepancy might be due to a difference in the distribution of speech sounds in the entire Ilocano language (estimated from this ~3000-word sample and shown in Table 3) vs. the distribution of speech sounds based only on two-syllable words (not estimated here or elsewhere). However, for many-to-most of the speech sounds, the actual and expected occurrences are very close (e.g., for /k/, 11 occurrences are expected in each list and the actual occurrences are 9, 9, 11, and 11). Overall, most of the differences between the expected and actual values are relatively small and are similar to those reported by others^{5,9,38}.

DISCUSSION AND FUTURE RESEARCH

As best as can be determined, this effort represents the first known attempt i) to estimate the distributions of speech sounds and syllable structures in Ilocano, and ii) to create word recognition lists for speech audiometry in the Ilocano language. Based on an analysis of 14,327 speech sounds from a language sample consisting of three magazine articles, four 50-item words lists in Ilocano (ILPB-50) were created. The Ilocano words in these lists 1) have a distribution of speech sounds that approximates the frequency of occurrence of speech sounds found in this language sample (i.e., these word lists are phonetically-balanced), 2) have two-syllables each, 3) have a distribution of syllable structures similar to that of the six most common structures for two-syllable words, and 4) were rated, by a majority of native Ilocano speakers, as 'common' words.

Ideally, word lists for speech audiometry should also meet several other criteria. Two very important criteria are interlist equivalency for listening and test-retest reliability (e.g., ^{34,41}). Unfortunately, we were not able to perform any perceptual studies or examine these factors

with the ILPB-50 word lists due to the extremely limited availability of native Ilocano speakers in the Saint Louis, Missouri area, where these analyses were done. Such behavioral tests should be performed as soon as possible. Additional suggestions for future word-list development include using a language sample based on spoken (not written) Ilocano, expanding the size of the language sample for estimating the distribution of speech sounds in Ilocano, examining the difference between the distribution of speech sounds for all words in Ilocano vs. that for two- syllable words in Ilocano, examining the distribution of consonant sounds as a function of their position in a word (i.e., initial-medial-final position considerations), and collecting data on the frequency of word usage in Ilocano^{31,40,42,43}.

APPENDIX A

Instructions:

Palawlawam ti maysa: Common / Somewhat Common / Uncommon

(For each word, please circle ONE: Common / Somewhat Common / Uncommon)

Common: Kanayon wenno assideg nga maaramat iti surat ken pannakisao iti ilocano.

(Common words are those used often in written and conversational Ilocano.)

Somewhat Common: Pagsasao nga saan unay nga maus-usar no agsurat ka ken makisarita ka iti ilocano.

(Somewhat Common words are those used often but not as often as Common words.)

Uncommon: Baliskas nga mausar sagpaminsan iti surat wenno pannakisarita iti ilocano. (Uncommon words are rarely used in written and conversational Ilocano.)

An English definition is provided (in parentheses) next to each word.

Example:

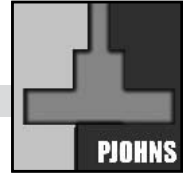
- BALSOG (fat)..... Common / Somewhat Common / Uncommon
LUKMEG (fat)..... Common / Somewhat Common / Uncommon

- 1A. ABAK (loss)..... Common / Somewhat Common / Uncommon
2A. ABAY (side-by-side)..... Common / Somewhat Common / Uncommon
3A. ABBONG (cover)..... Common / Somewhat Common / Uncommon
4A. ADAT (spicy)..... Common / Somewhat Common / Uncommon
5A. ADI (younger sibling)..... Common / Somewhat Common / Uncommon
6A. ADING (younger sibling)..... Common / Somewhat Common / Uncommon
7A. ADDA (to have)..... Common / Somewhat Common / Uncommon
8A. ADU (plenty)..... Common / Somewhat Common / Uncommon
9A. AGDAN (stairs)..... Common / Somewhat Common / Uncommon
10A. ALAS (ugly)..... Common / Somewhat Common / Uncommon
11A. ALUS (secondhand)..... Common / Somewhat Common / Uncommon
12A. AMIT (burden)..... Common / Somewhat Common / Uncommon
13A. ANAK (child)..... Common / Somewhat Common / Uncommon
14A. APAY (why)..... Common / Somewhat Common / Uncommon
15A. APO (grandparent)..... Common / Somewhat Common / Uncommon
16A. BABA (go down)..... Common / Somewhat Common / Uncommon
17A. BABOY (pig)..... Common / Somewhat Common / Uncommon
18A. BAEL (to carry)..... Common / Somewhat Common / Uncommon

APPENDIX B

Alphabetical lists of the ILPB-50 words
Ilocano word (English translation)

Table with 4 columns: List 1, List 2, List 3, List 4. Rows include words like abak (lose), abay (side-by-side), adi (younger sibling), ading (younger sibling), adat (spicy), adda (to have), agaw (to snatch), agong (nose), adu (plenty), agas (medicine), ala (to take), alas (ugly), agum (greed/envy), akar (move), anak (child), apay (why), apo (grandparent), amin (all), baboy (pig), bagas (uncooked rice), bael (to carry), baba (go down), baga (to tell), bagkat (to carry), bagay (well-matched), baeng (sneeze), bagio (storm), baket (old woman)



bagtit (crazy)	bagi (self)	baka (maybe)	balay (house)	lima (5)	libro (book)	ligay (wheel)	likud (back)
baklay (carry on shoulders)	bain (shame)	bakrang (side of body)	bannog (tired)	lubong (world)	linis (clean)	linong (shade)	lipat (forgetfulness)
bangon (rise)	baknang (rich)	bangles (spoiled)	baut (spank/whip)	magna (to walk)	lugan (car/vehicle)	luto (cook)	madi (don't like)
bantay (mountain)	banga (pot)	banio (bathroom)	bilang (count)	manen (once again)	malas (bad luck)	malem (afternoon)	malmes (to drown)
bingay (share)	bangsit (stink)	basol (fault/sin)	buksit (stomach)	ngato (high)	mureng (dirty clothes)	nagan (name)	nateng (vegetable)
daga (soil)	baro (new)	biit (short moment)	dadang (heat/warm)	padas (to try)	ninong (godfather)	nipis (cards)	pada (same)
dait (sew)	bigat (morning)	bugkaw (shout)	dagum (needle)	payso (true)	panaw (to leave)	paris (pair/same)	parit (prohibit)
damag (news)	binting (25 cents)	daan (old)	danum (water)	pilay (limp)	perdi (damage)	pidut (pick up)	pigsa (strong)
denggen (listen)	dagdag (to hurry)	dagsen (heavy)	dila (tongue)	punas (to wipe)	pili (choose)	pukaw (loss)	pulbos (powder)
dilnak (scald/burn)	dakes (bad)	dalan (road)	ganat (hurry)	ragsak (happiness)	pungan (pillow)	pungtot (anger)	rabaw (shallow)
dutdut (fur/hair)	damo (beginning)	danon (to arrive at)	gasat (luck/fortune)	ridaw (door)	raman (flavor)	ramas (to mix)	rang-ay (prosperity)
ganus (unripe)	denna (stay close)	diding (wall)	giddan (simultaneous)	rupa (face)	rikna (feeling)	riro (mistake)	rugit (dirty)
gasto (spend)	dugyot (dirty)	galis (slippery)	gusto (like)	sabon (soap)	saan (no)	saba (banana)	sabat (to meet)
gugot (gums)	gatas (milk)	garaw (unruly/wild)	ili (town)	sakit (sick)	sabong (flower)	sadut (lazy)	sagad (broom)
ibus (to finish up)	idi (before/past)	gaw-at (to reach)	ita (now)	siit (fish bone)	sugat (wound)	sida (food)	siding (mole)
ima (hand)	ipus (tail)	igid (edge)	kabsat (sibling)	suruk (excess/surplus)	surot (follow)	suli (corner)	supli (change [money])
itay (a short while ago)	kabaw (senile/forgetful)	isem (smile)	kali (to dig out)	tali (rope)	talna (calmness)	takaw (steal)	talaw (leave)
kadua (companion)	kagat (bite)	kabil (fight)	kelnat (to boil)	tinnag (to fall down)	tokar (music)	tamnay (tasteless)	tilmon (swallow)
kalman (yesterday)	kalub (to cover)	kama (bed)	kikit (little finger)	tulang (bone)	unay (very much)	ubing (child)	tulag (agreement)
kibin (hold hands)	kidem (close eyes)	kigtot (startle)	kurtar (cut hair)	uram (fire)	uray (wait/even though)	uneg (inside)	unnat (stretch)
lagto (jump)	lag-an (light/not heavy)	kugtar (kick)	lagip (memory)	waras (to distribute)	watwat (to exercise)	walin (to put aside)	walis (dust broom)
libas (to escape)	laing (smart)	lakay (old man)	lako (to sell)				

ACKNOWLEDGEMENTS:

We thank Dr. Helen Rumusud-Catcan (Marcos Memorial Hospital and Medical Center in Batac, Philippines) for background information on speech audiometry in the Philippines. Also, we thank Dr. Carl Rubino for his help with the linguistic description of Ilocano.

REFERENCES:

- Brandy, W.T. Speech Audiometry. in Handbook of Clinical Audiology (eds. Katz, J., Burkard, R.F. & Medwetsky, L.) 96-110 (Lippincott Williams & Wilkins, Philadelphia, 2002).
- Thibodeau, L.M. Speech Audiometry. in Audiology: Diagnosis, Vol. 1 (eds. Roeser, R.J., Valente, M. & Hosford-Dunn, H.) 281-309 (Thieme, New York, 2000).
- Jerger, J. & Jerger, S. Diagnostic significance of PB word functions. Archives of Otolaryngology 93, 573-580 (1971).
- Davis, H. Audiometry: Pure Tone and Simple Speech Tests. in Hearing and Deafness (eds. Davis, H. & Silverman, S.R.) 179-220 (Holt, Rinehart and Winston, New York, 1970).
- Lehiste, I. & Peterson, G.E. Linguistic Considerations in the Study of Speech Intelligibility. Journal of the Acoustical Society of America 31, 280-286 (1959).
- Carhart, R. Problems in the Measurement of Speech Discrimination. Arch Otolaryngol Head Neck Surg 82, 253-260 (1965).
- Ferrer, O. Speech audiometry: A discrimination test for Spanish language. Laryngoscope 70, 1541-1551 (1960).
- Liden, G. & Fant, G. Swedish word material for speech audiometry and articulation tests. Acta Oto-Laryngologica Supplement 114, 189-204 (1954).
- Turrini, M. et al. Bisyllabic words for speech audiometry: A new Italian material. Acta Otorhinolaryngologica Italica 13, 63-77 (1993).
- Harris, R.W., Goffi, M.V.S., Pedalini, M.E.B. & Gygi, M.A. Psychometrically equivalent Brazilian Portuguese bisyllabic word recognition spoken by male and female talkers. Pró-Fono Revista de Atualização Científica 13, 249-262 (2001).
- Nissen, S.L., Harris, R.W., Jennings, L.-J., Eggett, D.L. & Buck, H. Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. International Journal of Audiology 44, 379-390 (2005).
- Gordon, R.G., Jr. Ethnologue: Languages of the World. Fifteenth edn <<http://www.ethnologue.com>> (SIL International, Dallas, TX, 2005).
- Wikipedia contributors. Ilokano language -- Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org/w/index.php?title=Ilokano_language&oldid=78396709> (2006).
- Rubino, C.R.G. Ilocano: Ilocano-English/English-Ilocano Dictionary and Phrasebook, 266 (Hippocrene Books, New York, 1998).
- Carhart, R. Basic principles of speech audiometry. Acta Oto-Laryngologica 40, 62-71 (1951).
- Egan, J.P. Articulation Testing Methods. Laryngoscope 58, 955-991 (1948).
- Hirsh, I.J. et al. Development Of Materials For Speech Audiometry. The Journal of Speech and Hearing Disorders 17, 321-337 (1952).
- Kamm, C., Carterette, E.C., Morgan, D.E. & Dirks, D.D. Use of digitized speech materials in audiological research. Journal of Speech and Hearing Research 23, 709-721 (1980).
- Hood, J. & Poole, J. Influence of the speaker and other factors affecting speech intelligibility. Audiology 19, 434-455 (1980).
- Zubick, H.H. et al. Development of Speech-Audiometric Materials for Native Spanish-Speaking Adults. Audiology 22, 88-102 (1983).
- Magnusson, L. Reliable clinical determination of speech recognition scores using Swedish PB words in speech-weighted noise. Scandinavian Audiology 24, 217-223 (1995).
- Denes, P.B. On the Statistics of Spoken English. The Journal of the Acoustical Society of America 35, 892-904 (1963).
- French, N.R., Charles W. Carter, J. & Walter Koenig, J. The Words and Sounds of Telephone Conversations. Bell System Technical Journal 9, 290-324 (1930).
- Martin, F.N., Champlin, C.A. & Perez, D.D. The Question of Phonetic Balance in Word Recognition Testing. Journal of the American Academy of Audiology 11, 489-493 (2000).
- Weisleder, P. & Hodgson, W.R. Evaluation of Four Spanish Word-Recognition-Ability Lists. Ear and Hearing 10, 387-393 (1989).
- Harris, R.W., Goffi, M.V.S., Pedalini, M.E.B., Gygi, M.A. & Merrill, A. Psychometrically equivalent trisyllabic Brazilian words spoken by male and female speakers of the Brazilian Portuguese. Pró-Fono Revista de Atualização Científica 13, 37-53 (2001).
- Owens, E. Intelligibility of words varying in familiarity. Journal of Speech and Hearing Research 4, 113-129 (1961).
- Schultz, M.C. Word familiarity influences in speech discrimination. Journal of Speech and Hearing Research 7, 395-400 (1964).
- Epstein, A., Giolas, T.G. & Owens, E. Familiarity and intelligibility of monosyllabic word lists. Journal of Speech and Hearing Research 11, 435-438 (1968).
- Rosenzweig, M.R. & Postman, L. Intelligibility as a function of frequency of usage. Journal of Experimental Psychology 54, 412-422 (1957).
- Rosenzweig, M.R. & Postman, L. Frequency of Usage and the Perception of Words. Science 127, 263-266 (1958).
- Egan, J.P. Remarks on Rare PB Words. Journal of the Acoustical Society of America 29, 751 (1957).
- Pisoni, D.B. & Luce, P.A. Recognizing spoken words: The Neighborhood Activation Model. Ear and Hearing 19, 1-36 (1998).
- Kirk, K.I., Eisenberg, L.S., Martinez, A.S. & Hay-McCutcheon, M. Lexical Neighborhood Test: Test-Retest Reliability and Interlist Equivalency. Journal of the American Academy of Audiology 10, 113-123 (1999).
- Kirk, K.I., Pisoni, D.B. & Osberger, M.J. Lexical Effects on Spoken Word Recognition by Pediatric Cochlear Implant Users. Ear & Hearing 16, 470-481 (1995).
- Eberling, C., Ludvigsen, C. & Lyregaard, P.E. DANTALE: A new Danish speech material. Scandinavian Audiology 18, 169-175 (1989).
- Rubino, C.R.G. Ilocano Dictionary and Grammar: Ilocano-English, English-Ilocano, 863 (University of Hawaii Press, Honolulu, 2001).
- Berruecos, P. & Rodriguez, J.L. Determination of the phonetic percent in the Spanish language spoken in Mexico City, and formation of PB. lists of trochaic words. International Journal of Audiology 6, 211-216 (1967).
- Tillman, T.W. & Carhart, R. An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6. 1-12 (USAF School of Aerospace Medicine, Brooks Air Force Base, TX, 1966).
- Peterson, G.E. & Lehiste, I. Revised CNC Lists for Auditory Tests. Journal of Speech and Hearing Disorders 27, 62-70 (1962).
- Skinner, M.W. et al. Evaluation of Equivalency in Two Recordings of Monosyllabic Words. Journal of the American Academy of Audiology 17, 350-366 (2006).
- Thorndike, E.L. & Lorge, I. The Teacher's Word Book of 30,000 Words, (Bureau of Publications, Teachers College, Columbia University, New York, 1944).
- Harris, J. WorldCount. <<http://www.wordcount.org/main.php>> (2003).