AUDIOLOGY · 청능재활 RESEARCH PAPER
2012;8:183-189

ISSN 1738-9399

무오류 음절을 이용한 노인성 난청의 언어인지검사

한림대학교 자연과학대학 언어청각학부[1] · 한림대학교 청각언어연구소[2] · 일리노이대학교 전자컴퓨터공학과[3]

한 우 재[1,2] · Jont B. Allen[3]

ABSTRACT

Speech Perception Test of Zero-Error Consonant-Vowel Syllables in Presbycusis

Woojae Han[1,2] and Jont B. Allen[3]

[1]*Div. of Speech Pathology and Audiology, College of Natural Science, Hallym University, Chuncheon, Korea*
[2]*Research Institute of Audiology and Speech Pathology, Hallym University, Chuncheon, Korea*
[3]*Dept. of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, IL, USA*

The purpose of this study is to measure speech perception ability in presbycusis, using zero-error consonant-vowel (CV) syllables as a function of signal-to-noise ratio (SNR). Seventeen elder people with symmetric mild-to-moderate sensorineural hearing loss were tested with 14 syllables spoken by two different speakers at four different SNRs. Error rate and entropy calculation were used to analyze data. The results show that both error rate and entropy significantly increased as noise increased. In particular, /ba, va, ga, na/ were affected by noise level in both error rate and entropy, yet /za, ʒa/ were only in entropy. We conclude that evaluating speech perception ability using zero-error CV syllables might scrutinize the unique characteristics of presbycusis. We also find that entropy analysis is very useful in even including response's uncertainty which is not found in the typical error rate analysis.

KEY WORDS : Consonant perception, Speech perception, Error rate, Entropy, Signal-to-noise ratio, Hearing-impaired listener

INTRODUCTION

Hearing-impaired listeners and hearing aid users commonly claim that it is difficult to communicate in noisy environments. This problem has been known clinically for a long time and has been documented by numerous researchers on speech perception (Zurek & Delhorne, 1987). Moreover, deficits in speech perception associated with hearing loss represent a major component of elderly persons' communication disability (Humes & Roberts, 1990). The most obvious and well-documented peripheral deficit in the elderly is the presence of high-frequency sensorineural hearing loss and poor speech perception scores (Humes & Roberts, 1990). Also, the central problem in the elderly population affects speech perception scores (Gates & Mills, 2005). However, the researchers have been concerned about the poor prediction of speech reception threshold (SRT) and word recognition score (WRS) from audiometric data. They also have a negative opinion about the comparability of various clinical speech tests with different types of hearing-impaired listeners and different testing conditions. Although such test scores are useful for estimating the extent to which listeners are handicapped, they do little to illuminate the nature of the speech perception problem (Bilger & Wang, 1976).

In audiology clinics, the most popular speech perception tests include SRT, WRS, and the Hearing in Noise Test (HINT). Plomp's SRT measures threshold using spondee words, whereas WRS and HINT measure speech compre-

논문접수일 : 2012년 10월 17일
논문수정일 : 2012년 12월 03일
게재확정일 : 2012년 12월 07일
Corresponding Author : Woojae Han, 200-702 Hallym Univ., Div. of Speech Pathology and Audiology, Hallymdaehakgil 1, Chunchon, Korea
Phone number : (033) 248-2216
E-mail : woojaehan@hallym.ac.kr

hension at the level of supra-threshold, using monosyllablesand sentences, respectively. We agree that all three tests have been popular and fast in the clinic. However, none of them has not strict normal hearing criteria, although Duncan and Aarts (2006) insist that HINT is much better in overcoming shortcomings associated with the typical speech test battery used to measure hearing handicaps such as poor representation of natural speech spoken by a spondee (or SRT) and the floor and ceiling effects associated with percent correct scoring (or WRS). That is, there is no difference between 0 dB HL and 20 dB HL, or between 0% error and 20% error in terms of the definition of normal hearing. In this context, the consonant-vowel (CV) syllable test has an advantage and is a more critical measurementofthe hearing-impaired listener's perceptual ability (Turner et al., 1992). For instance, Tolhurst (1949) measured the threshold of detectability for individual voiceless consonants, which resulted in the recognition of each consonant up toa 30-dB range spanned even in normalhearing listeners. Later, Kewley-Port (1991) measured the detection thresholds of isolated vowels in normal hearing listeners, and found a 22-dB range across the vowel tokens. Hence, these large differences should be considered clinically in hearing-impaired as well as normal hearing listeners to more accuratelymeasure their perceptual ability (Han, 2012).

The Articulation Index (IA) originally described by French and Steinberg (1947) and Fletcher and Galt (1950) has proven to provide an accurate prediction of speech perception performance. The success of this index as a predictor of speech recognition performance, however, confirms the primary importance of speech spectrum audibility in explaining the speech recognition deficits of only for most of the young hearing-impaired population (Humes & Roberts, 1990).

On the other hand, Singh proposed zero-error utterance (Singh, 2010). For instance, when using utterances for which normal hearing listeners incur no errors (i.e., about 90% correct or more ≥ -2 dB SNR), called 'zero errors', results of the tests of hearing-impaired listeners could be much more confidential in representing their unique problem of speech perception. In other words, it is important to specify the comparative reference when stating that hearing-impaired listeners have additional difficulty listening to speech in noise because, of course, normal hearing listeners also have difficulty understanding speech in noise (Zurek & Delhorne, 1987). To control the normal hearing listeners, Zurek and Delhorne employed spectrally shaped masking noise that raised the detection thresholds of the normal-hearing listeners to match those of the hearing-impaired listeners. The noise-masked normal listeners were then given speech tests under conditions identical to those experienced by impaired listeners (Zurek & Delhorne, 1987). However, their experiment had a shortcoming in which the auditory plasticity of hearing-impaired listeners was not considered.

The purpose of the current study was to estimate common error patterns that often occur in aging hearing-impairment, especially presbycusis, as a function of SNR while presenting only zero-error CV syllables. A second purpose was to identify their uncertainty when listening to the syllables, while applying entropy calculation.

MATERIALS AND METHODS

Subjects

Seventeen subjects (eight males and nine females) were recruited from the Urbana-Champaign community. All subjects were native speakers of American English and ranged in age from 65 to 84 years (mean = 73.23 years, SD = 6.95). They had normal middle-ear status (A type of tympanogram) and symmetric mild-to-moderate sensorineural at PTA (pure tone average; average in hearing threshold at 0.5, 1, 2, and 4 kHz). The criteria of symmetric hearing loss were less than 5-dB HL difference between two ears (Table 1). Informed consent was obtained and approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign (UIUC IRB protocol #05149).

Stimuli

Isolated English consonant-vowel (CV) syllables were chosen from the Linguistic Data Consortium (LDC) 2205S22 database, spoken by 18 native speakers of American English (Fousek et al., 2004). The CV syllables consisted of 14 consonants (six stops, six fricatives, and two nasals, i.e., /p, b, t, d, k, g, f, v, s, ʃ, z, ʒ, m, n/) followed by the /a/ vowel. Two fricatives, /θ/ and /ð/, were not used

Table 1. Subject's information and results of the hearing screening tests

Subject	Age	Pure-tone thresholds (dB HL in Hz)										SRT	WRS
		125	250	500	1000	1500	2000	3000	4000	6000	8000	(dB HL)	(%)
1	82	40	40	45	45	45	45	45	45	55	50	45	70
2	81	45	45	45	50	45	45	55	55	55	55	45	75
3	65	40	40	45	45	50	45	55	35	30	25	40	80
4	68	30	30	25	30	25	35	55	50	50	50	30	80
5	66	25	25	25	25	25	30	55	50	55	50	25	85
6	74	30	30	30	30	30	45	40	45	55	55	35	80
7	74	30	30	30	20	25	30	45	50	55	50	30	75
8	84	40	35	35	25	30	35	35	50	55	50	30	80
9	83	30	25	20	25	30	40	45	55	55	50	25	80
10	72	20	25	20	30	30	35	35	35	40	45	25	70
11	70	25	25	25	25	30	35	45	50	40	55	30	80
12	79	20	20	25	20	30	20	35	50	45	55	25	75
13	80	20	20	25	25	30	30	40	35	45	50	25	80
14	65	20	20	20	20	20	20	35	55	45	40	20	80
15	65	20	20	20	25	25	20	30	45	40	30	20	85
16	70	20	25	20	25	25	20	40	55	35	40	20	85
17	67	20	25	25	20	25	25	25	40	35	40	25	80

in the experiment, as they have high error, even for normal hearing listeners (Li et al., 2010; Phatak & Allen, 2007). To reduce the time of administration, only two talkers (one male and one female) were selected per syllable. In total, there were 14 × 2 = 28 different utterances. All 28 utterances had zero error for SNR ≥ -2 dB (i.e., about 90% correct or more ≥ -2 dB SNR) across the 14 normalhearing listeners in the previous study (Phatak & Allen, 2007).

Experimental Procedure

All of the subjects had one practice session, with 14 syllables (i.e., same CV set consisting of different speakers) in quiet, before they began the experiment, employing a 14-alternative forced-choice response task. These 14 syllables were always different from the practice syllables to minimize learning effects. Syllable presentation was randomized over consonants and speakers. Three SNRs, 12, 6, and 0 dB, and quiet conditions were tested in order to establish an easy listeningcondition. Through an inserted earphone (ER2), only one ear (right or preferred ear) was tested.

The experiment consisted of two sessions. In the first session, each of the 28 utterances was presented four times at each SNR. This resulted in 28 utterances × 4 SNRs × 4 presentations = 448 trials. For each utterance at each SNR, the correct percentage score was calculated. The possible scores were 0% (0/4), 25% (1/4), 50% (2/4), 75% (3/4), and 100% (4/4). In the second session, the number of trials depended on the subject's performance in the first session. Across the two sessions each utterance was presented between 5 and 10 times, depending on the error rate in the first session, and therefore each consonant was presented between 10 and 20 times at each SNR. The rationale behind this experimental design was to increase the sample size as a function of the score and to obtain more data when more errors were being made. The total number of trials per syllable (sum of sessions I and II) was not same for all subjects. About 800-1,000 trials were presented to each subject and the experiment took a total of 30-40 minutes per subject.

Data Analysis

To analyze the data, repeated measure analysis of variance (ANOVA) and the Bonferroni post hoc test were used. Error rate simply referred to percent error as a function of SNR per syllable. However, to our knowledge, no study has examined entropy calculation in the audiology field. Perception scores, when expressed as simple percent correct for individual syllables in a closed-set task, can be influenced by response bias of the subjects. For example, a subject participating in a closed-set response task might consistently respond with a single answer, thus leading to a result where one syllable would be identified at perfect levels regardless of its presentation level (Turner et al., 1992). Thus, we introduced entropy as a function of SNR per syllable like an error rate. Entropy is a measure of the uncertainty associated with a random variable and of the average information the subject is missing when he/she does not know the value of the random variable (Shannon, 1948). It quantifies the expected value of the information contained in a message, usually in units such as bits (see the Formula 1, which we used in this paper). Thus, high entropy is not equal to high error rate. In other words, even though two stimuli, /ba/ and /pa/, show same 80% error rate in a particular listener, when /ba/ has 1 bit and /pa/ has 1.5 bits in the entropy of the listener, we can say that his/her uncertainty of /ba/ is lower than /pa/, which means /ba/ has more consistent error patterns than /pa/.

$$H = -\sum_i p(x_i)\log_2 p(x_i)$$ -- <Formula 1>

RESULTS

The percentage of errors increased significantly as the noise increased [F (3, 48) = 40.211, $p < 0.000$]. The error was the least at quiet condition (mean = 9.90%, SD = 3.111). The Bonferroni post-hoc test resulted in a significant difference between 12 dB (mean = 12.06%, SD = 2.917) and 6 dB (mean = 19.01%, SD = 3.668), and between 6 dB and 0 dB (mean = 29.32%, SD = 4.323). In addition, all consonants showed significant difference as SNR, the subjects had significant difference across the syllables [F (13, 208) = 4.596, $p < 0.000$]. In general, /ta/ had the lowest error (mean = 3.87%; SD = 2.00) and /da, ʃa/ were next lowest error (mean = 4.19%, SD = 2.29; mean = 4.39%, SD = 2.81, respectively), while /ʒa, va, ba, za, ga/ had the highest errors (mean = 32.94%, SD = 9.75; mean = 28.24%, SD = 5.36; mean = 27.80%, SD = 4.21; mean = 27.36%, SD = 6.36; mean = 24.31%, SD = 6.29, respectively) (Fig. 1). There is an interaction effect between SNR and the consonants [F (39, 624) = 3.435, $p < 0.000$]. That is, /ta, da, ka, ʃa, ma/ were interacted among 0, -6, and -12 dB SNR.

The entropy was also statistically analyzed with repeated measure ANOVA to find significantpredictors. As noise increased, the entropy significantly increased [F (3, 48) = 78.045, $p < .000$]. That is, the quiet condition (mean = .276; SD = .074) was significantly lower entropy than 12 dB (mean = .390; SD = .071), 12 dB was lower than 6 dB (mean = .585; SD = .085), which was much lower than 0 dB (mean = 1.086; SD = .111). Subjects had significant difference in entropy of consonant perception [F (13, 208) = 11.187, $p < .000$]. /ʃa/ had the lowest entropy (mean = 0.188; SD = .103) and /da, ta/ were the next lowest error, while /va, ba, za, ga, ʒa/ had the highest entropy (mean = 1.111, SD = .131; mean = 1.033, SD = .116; mean = .815, SD = .144; mean = .758, SD = .115; mean = .733, SD = .142, respectively) (Fig. 2). There is an interaction effect between SNR and consonants [F (39, 624) = 7.166, $p < .000$]. CVs /ta, da, ʃa/ were interacted among 0, -6, and -12 dB SNR.

Fig. 3 shows the relationship of error rate and entropy. Both error rate and entropy were generally increased when the noise level was changed from quiet to 0 dB SNR. However, the level of noise might affect entropy, rather than error rate. Although the error rate was not different, while ranging from 0 to 30% at either quiet or 12 dB SNR condition, the entropy was increased up to 1 at only 12 dB SNR. At 0 dB SNR, /ba, ga, va/ showed entropy as high as 2.

DISCUSSIONS

Although it is acknowledged that sensorineural hearing loss in presbycusis is generally accompanied by a loss of speech perception ability, relatively little research has been carried out documenting the nature of speech perception loss (Bilger & Wang, 1976).

The present study documents the speech perception of presbycusis (i.e., aging hearing impairment) when presenting only zero-error CV syllables, in terms of error rate and entropy analysis. Error rate means amount of misperceived syllables, while entropy is uncertainty in selecting what the listeners heard, regardless of either correct or incorrect answer. We found that as noise increased, error rate also increased, regardless of the kind of syllables. Frisina & Frisina (1997) proposed that when cognitive functioning was not affected, dysfunction of speech recognition in noise usually remained in elder people. This could extend auditory cortex temporal-resolution dysfunctions in accounting for the

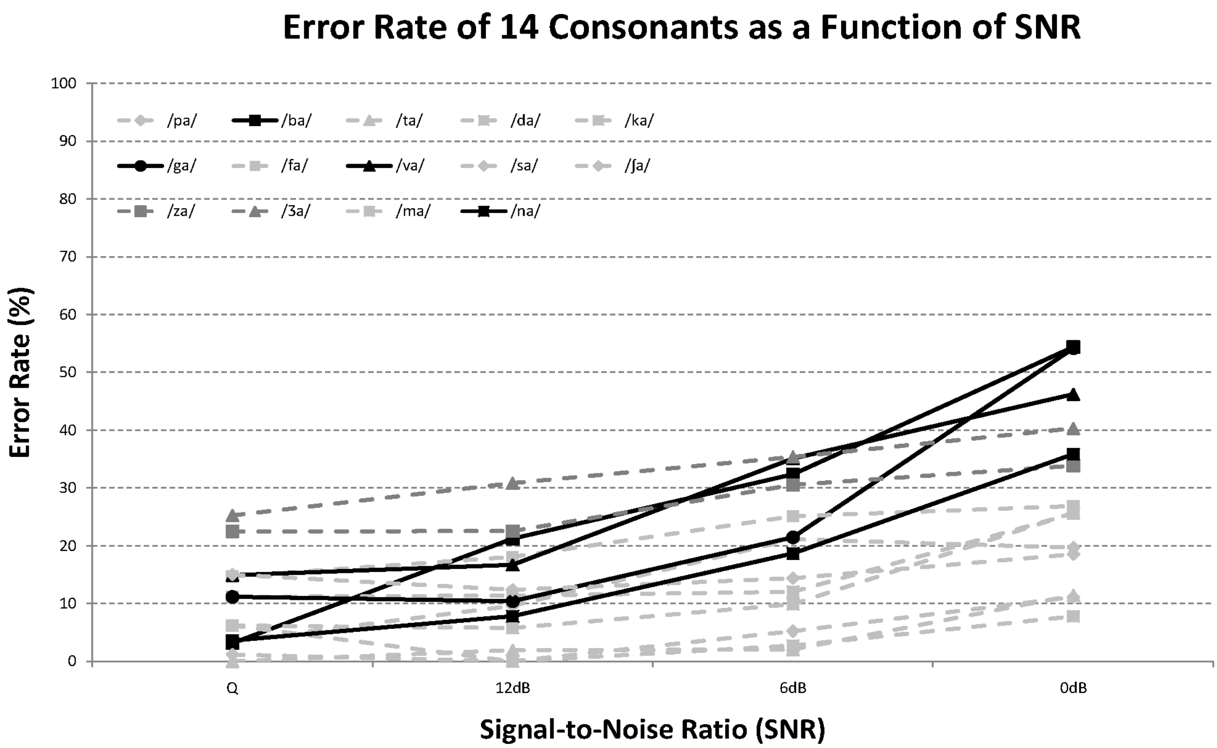

Figure 1. Error rate of 14 consonants as a function of SNR. Each curve refers to each syllable. Solid curves mean syllables sensitive to the noise level. For example, /ba, ga, va, na/ were much affected as the noise increased.

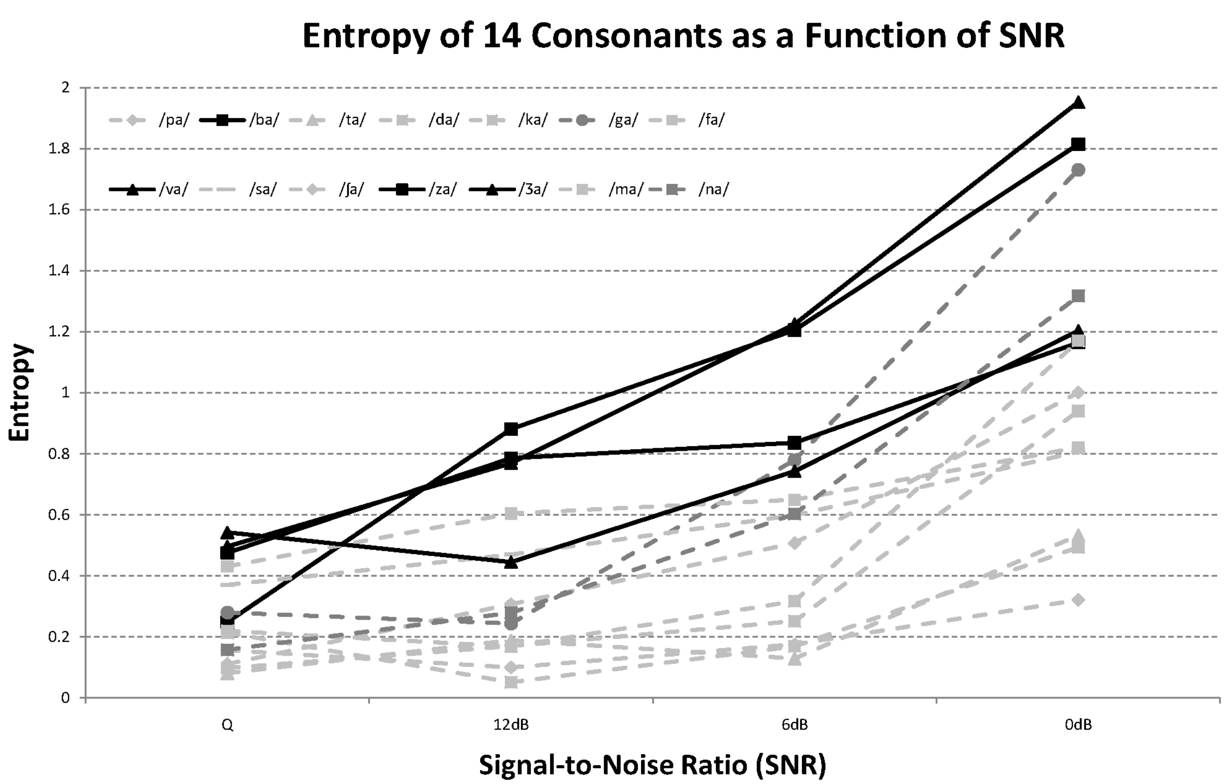

Figure 2. Entropy of 14 consonants as a function of SNR. Four solid curves and two dark dashed curves of 14 curves include syllables sensitive to the noise level. For instance, /ba, va, ga, na, za, ʒa/ were dramatically changed in entropy as noise increased.

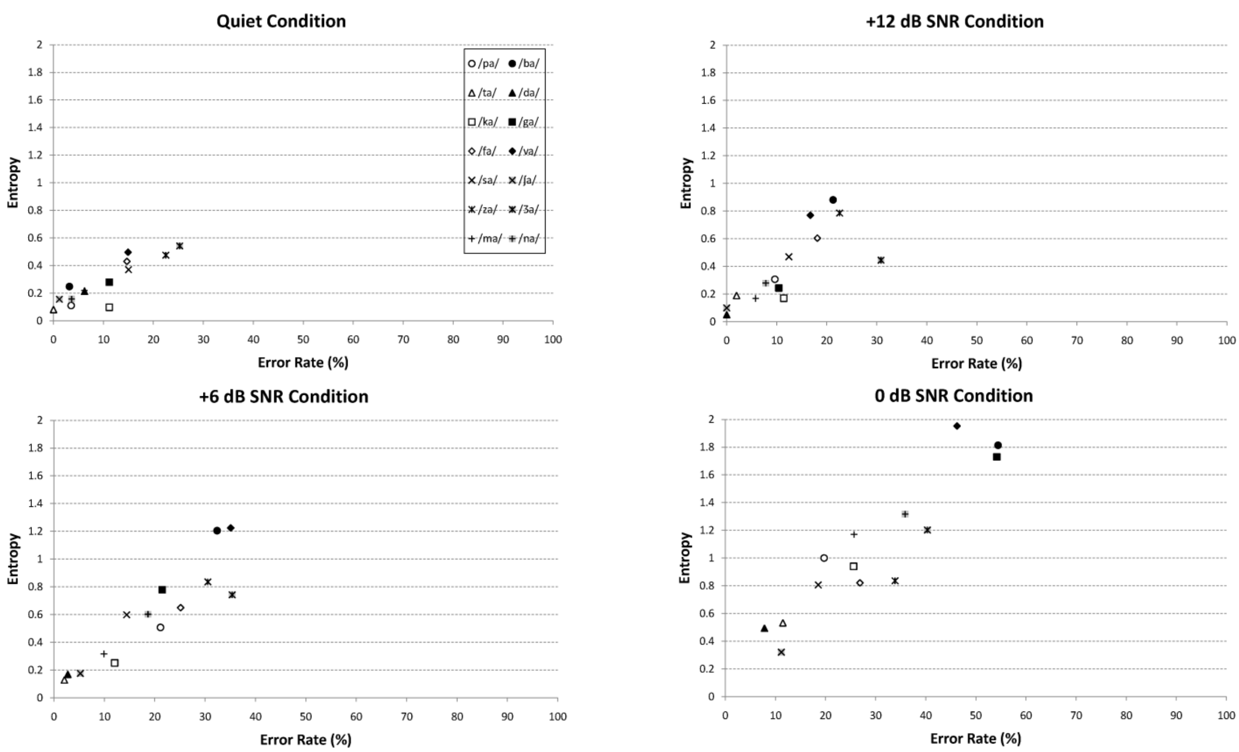

Figure 3. Consonant perception cluster of error rate (x-axis) and entropy (y-axis). Four panels represent each SNR. As the noise increased, there was less cluster, especially at 0 dB SNR.

observed differences in speech processing.In our findings, some syllables, i.e., /ba, ga, va, na/, were very sensitive to noise. This might explain that those syllables having a little longer time relative to release than others are not easy to be perceived in presbycusis under noise condition (Gordon-Salant& Fitzgibbon, 1993). Gordon-Salant et al.(2007) also support our findings in that these difficulties are thought to reflect underlying problems in auditory temporal processing by elder people that may occur at various stages of peripheral and central auditory processing or as a result of age-related cognitive slowing.

Entropy also changed a lot as noise increased. Compared to the error rate, entropy was much affected in noise, which means presbycusis has large uncertainty in perceiving syllables in a higher noise situation. Davis & Johnsrude (2007) propose evidence foran interactive processing system in which bottom-up andtop-down processes combine to support speech perception. That is, this interactive account provides mechanisms by whichperceptual processing can rapidly change so as to perceive (i.e., bottom-up) and comprehend (i.e., top-down) speech. We could interpret that bottom-up processing is for perceiving the syllable correctly or incorrectly, and top-down processing is for comprehending the syllable, and thus choosing the correct answer during an experiment having certainty. Although the two key processes could be interactive, the researchers conclude that the top-downinteractive mechanisms within auditory networks play a key role in explaining the perception of spoken language. In other words, to see certainty (or uncertainty) is necessary for successful speech perception diagnosis in presbycusis. The zero-error syllables test and entropy analysis would need to be conducted in order to generalize the findings to these difficult everyday listening conditions of presbycusis through following studies of test-retest reliability and faster method for the clinical usage.

Acknowledgement

This research was supported by the National Institutes of Health (NIH) Grant (RDC 009277A). We thank Riya Singh and several graduate students in UIUC HSR laboratory for

critical comments on experimental design of the paper.

REFERENCES

Bilger, R. C. & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *19*, 718-748.

Davis, M. H. & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Research*, *229,* 132-147.

Duncan, K. R. & Aarts, N. L. (2006). A comparison of the HINT and Quick SIN tests. *Journal of Speech-Language Pathology and Audiology*, *30(2)*, 86-94.

Fletcher, H. & Galt, R. H. (1950). The perception of speech and its relation to telephony. *Journal of the Acoustical Society of America*, *22*, 89-151.

Fousek, P., Svojanovsky, P., Grezl, F., & Hermansky, H. (2004). New nonsensesyllables database - Analyses and preliminary ASR experiments. In *The InternationalConference on Spoken Language Processing (ICSLP)*.

French, N. R. & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19,* 90-119.

Frisina, D. R. & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research*, *106,* 95-104.

Gates, G. A. & Mills, J. H. (2005). Presbycusis. *The Lancet*, *336,* 1111-1120.

Gordon-Salant, S. & Fitzgibbon, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research*, *36*, 1276-1285.

Gordon-Salant, S., Fitzgibbon, P. J., & Friedman, S. A. (2007). Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. *Journal of Speech and Hearing Research*, *50*, 1181-1193.

Han, W. (2012). Case study of speech perception in presbycusis, *Audiology*, *8*, 116-122.

Humes, L. & Roberts, L. (1990). Speech-recognition difficulties of the hearing impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research*, 33, 726-735.

Kewley-Port, D. (1991). Detection thresholds for isolated vowels. *Journal of the Acoustical Society of America, 89(2)*, 820-829.

Li, F., Menon, A., & Allen, J. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *Journal of the Acoustical Society of America, 127(4),* 2599–2610.

Phatak, S. & Allen, J. (2007). Consonant and vowel confusions in speech-weighted noise. *Journal of the Acoustical Society of America, 121(4),* 2312–2326.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27(3),* 379-423.

Singh, R. O. (2010). A theory of speech perception in normal and hearing-impaired ears. (Master thesis) University of Illinois at Urbana-Champaign, IL, USA.

Tolhurst, G. C. (1949). Audibility of the voiceless consonants as a function of intensity. *Journal of Speech and Hearing Disorders, 14*, 210-215.

Turner, C. W., Fabry, D. A., Barrett, S., & Horwitz, A. R. (1992). Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research*, *35*, 942-949.

Zurek, P. M. & Delhorne, L. A. (1987). Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairement. *Journal of the Acoustical Society of America, 82(5),* 1548-1559.