Auditory-Visual Augmentation of Thai Lexical Tone Perception in the Elderly

Benjawan Kasisopa[1], Sudaporn Luksaneeyanawin[2], Suparak Techacharoenrungrueang[2], Denis Burnham[1]

[1]MARCS Institute, University of Western Sydney; [2]Center for Research in Speech and Language Processing, Chulalongkorn University
b.kasisopa@uws.edu.au, sudaporn.l@chula.ac.th, suparak.t@student.chula.ac.th, d.burnham@uws.edu.au

ABSTRACT

This study investigated the effects of aging, auditory and auditory-visual perception of lexical tone of native Thai listeners. Elderly and younger Adults' discrimination of the 5 Thai tones was investigated in audio-visual (AV), audio-only (AO), and visual-only (VO) conditions at two inter-stimulus intervals (ISIs) [500 and 1500 ms] in clear and noisy conditions. Generally, the Elderly performed more poorly than the Adults, but in both groups there was similar ranking of the relative discriminability of tone contrasts. Notably, in noise there was better tone discrimination in AV than in AO, and this was equally the case for young Adults and the Elderly. This shows that the elderly can and do use visual speech information to augment auditory perception of tone. The Elderly also benefitted more from the 1500 ms ISI suggesting that their tone perception is better when a more phonological (rather than acoustic) mode of processing is available.

Keywords: tone perception, auditory-visual, aging, discrimination, speech in noise

1. INTRODUCTION

Lexical tones are linguistically contrastive in 70% of the world's languages [1]. While duration, F2 values, and amplitude (perceived as vowel length and vowel height, and loudness respectively) all play some part in lexical tone [2, 3, 4], the principal feature of lexical tone is F0 (perceived as pitch). In tone languages, a change in F0 height or contour can result in a change in the identity of the lexical tone and the meaning of the word. For example, Bangkok Thai has 5 tones, three level or static (Low-21, Mid-33, and High-45) and two contour or dynamic (Falling-241 and Rising-315) (numbers are Chao values [5]). For example, [maa33] (มา)-mid tone means 'to come', [maa21] (หม่า)-low means 'brew', [maa45] (ม้า)-high means 'horse', [maa241] (ม่า)-falling means 'grandmother', and [maa315] (หมา)-rising means 'dog'.

A change in the identity of consonants and/or vowels also changes the meaning of a word. There has been extensive investigation of the auditory perception of consonants and vowels [6], and a number of factors have been found to influence speech discrimination. Of interest here, firstly, auditory speech perception has been found to be affected by the inter-stimulus interval (ISI) between pairs of sounds presented in a same-different task. At a shorter ISI (500ms), listeners show evidence of within-category auditory-level discrimination, whereas at a longer ISI (1500ms), only phonemic categorisation is evident [6, 7]. Secondly, it is now very clear that visual speech information is used in speech perception when it is available [8], even in undegraded listening conditions [9].

While auditory [6] and auditory-visual speech perception [10] and ISI [6, 7] research has predominantly focussed on consonants and vowels, of late there has been a welcome increase in perception studies involving lexical tone. However, these tone perception studies have focussed on young adult populations; and we know that speech perception declines with age, but little is known about the specific effects on elderly listeners. The study reported here extends these studies to auditory and auditory-visual perception of Thai lexical tone in elderly native Thai listeners.

1.1. Aging and speech perception

It has long been known that the elderly have more difficulty understanding speech than young listeners, especially in background noise [11]. Increased age is usually accompanied by frequency-range degradation that begins at the high end of the usual 20 Hz to 20 kHz range of human hearing. The gradual high-frequency loss eventually encroaches upon the range necessary for speech perception, generally regarded as being from 250 to 4000 Hz [12]. Language in old age has been an active research area since early investigations in cognitive aging in the 1960s [13] through to the present [14], because of the importance of language throughout the life span not only in cognitive function, but also in social interaction. Decline in language processing, such as increased difficulty in understanding spoken language, undermines older adults' ability and desire to communicate, and can erode their own and others' evaluation of their language competence [15, 16].

As yet, there are no studies of the effect of aging on lexical tone perception, but the results of studies on effects of aging on other aspects of speech perception in noise have found that elderly speech perception decreases as age increases and signal-to-noise ratio (SNR) decreases [11, 15]. In addition to such direct effects of a degraded signal [11, 17], there is also evidence for indirect effects of perceptual decline in older adults; according to a resource capacity account greater difficulty in perceptual processing affects subsequent processes by constraining available resources [15].

1.2. Auditory Perception of Tone

Cross-language studies have shown that native listeners of Thai perceive tones better than do non-native listeners of other tone or pitch-accent languages, who in turn perceive tones better than non-native non-tone language listeners [18, 19]. There are also consistent differences in discrimination of particular pairs of tone types, with better discrimination of Dynamic-Dynamic (DD) pairs (Falling-Rising/FR) than Static-Static (SS) (Mid-Low/ML, Mid-High/MH, Low-High/LH) and in turn better for Static-Dynamic (SD) pairs (Low-Rising/LR, Mid-Rising/MR, High-Rising/HR, Low-Falling/LF, Mid-Falling/MF, High-Falling/HF) [19]. Further, in accord with the proposed physiological bias towards better perception of rising pitch contours based on studies of the brainstem frequency following response (FFR) in the [20], it has been found that within SD pairs, those involving rising tones are discriminated better than those involving falling tones [19].

1.3 Auditory-Visual Perception of Tones

A 40%–80% augmentation of auditory-only (AO) speech perception has been found when speech in noise is accompanied by the speaker's face [21], and there is now evidence for visual influences in lexical tone perception in Cantonese [22], Mandarin [25] and Thai [19]. While this auditory-visual augmentation of lexical tone perception is evident across native tone-, non-native tone- and non-native non-tone language listeners [19, 23, 24], there is, somewhat paradoxically, better use of visual information for tone in visual only (VO) conditions by non-native, non-tone language listeners than by native or non-native tone language listeners [19, 25].

In this study the auditory and auditory-visual tone perception of tone is investigated in elderly Thai listeners, and compared with a reference group of adult Thai listeners. Based on the above literature, the hypotheses are:

a) Younger will perform better than older listeners in auditory tone perception, especially in noise.
b) Given that elderly listeners may have reduced attention to auditory input [11, 15, 17], the elderly will be able to make better use of visual information than the young adults (i) in VO and (ii) for auditory-visual augmentation (AV vs AO) of tone perception in noise.
c) Young adults and the elderly will perform similarly on the discrimination of DD, SD and SS tone pairs, though as there is little research in this area, exact predictions are not possible.

2. METHOD

2.1. Participants

Thai Elderly: 37 native Thai elderly listeners (mean age 63 years, *SD*=3.1, 24 females) were recruited in Bangkok, Thailand. Most had normal hearing with pure-tone thresholds lower than 25 dB HL; a few had mild hearing loss with thresholds lower than 40 dB HL at octave frequencies from 250 Hz to 8 kHz in both ears.

Thai Adults: 36 native Thai adult listeners (mean age 29 years, *SD*=4.0, 21 females) were recruited in Sydney, Australia. Mean duration in Australia at testing was 2 years (*SD*=2.8). All had normal hearing with pure-tone thresholds lower than 25 dB HL in both ears.

2.2. Experimental design

A 2 [age groups: adults/elderly] x 2 [ISI: 500/1500 ms] x (2 [clear/noise conditions] x 3 [modes: AV/AO/VO] x 10 [tone contrasts – SS (ML/MH/ LH); SD (MF/HF/LF/MR/HR/FR); and DD (FR)] x 4 [order of same/different pairing conditions: 2 different, AB & BA, trials, and 2 same, AA & BB, trials] x 2 [repetitions for Adults] or x 1 [repetition for Elderly]).

2.3. Stimulus materials

Stimuli consisted of 6 Thai CV syllables (C = [k/k^h]; V = [a:/i:/u:]) each carrying all 5 Bangkok Thai tones. These are either words (n = 21) or nonwords (n = 9). The 30 syllables were recorded in citation form from a native Thai female speaker. Productions were recorded audio-visually in a sound-treated booth with a 25 fps, 720 x 576 pixels, and 48 kHz 16-bit audio. Three good quality exemplars of each syllable were selected. Sound level was normalised and all videos were compressed using msmpeg4v2 codec. In noise conditions, a multi-talker Thai speech babble track was played with each stimulus, with -8 dB SNR. Note, the VO mode also contained babble noise in the noise condition and a still image of the speaker was displayed in the AO mode.

2.4. Procedure

Participants were tested individually in a sound-attenuated room on individual laptop computers running DMDX software [26] with the visual component presented in the centre of the screen. The auditory stimuli were presented via Sennheiser HD 25-1 II headphones connected through an EDIROL /Cakewalk UA-25EX USB audio interface unit at a comfortable hearing level (M=65dB).

For the young adults, there were 480 trials split into 2 test files (240 trials in each) for blocked testing of clear and noise stimuli. The clear and noise blocks were each split into two 120-trial blocks. In each block, 40 trials in each mode made up of 10 tone pairs in 4 AB orders were presented randomly and across blocks different stimulus exemplar repetitions were used. Block order was counterbalanced between participants. At the start of each block, 4 extra trials were presented: 1 AV, 1 AO, and 1 VO trial in the training session, then another AV trial placed at the start of the test session as a warm-up trial. Half the numbers of trials were used for each block (120 trials) for the elderly. All participants were asked to determine whether the two tones were the same or different with the time-out limit for each trial at 5 seconds. If a participant failed to respond on a particular trial, one additional chance to respond was given in an immediate repetition of that trial. Participants were given breaks between each block.

2.5. Data processing and analysis

d' scores were calculated for each of the 10 tone pairs in each condition ($d' = Z(\text{hit rate}) - Z(\text{false positive rate})$ with adjustments made for probabilities of 0 (=.05) and 1 (=.95), where *hit* is a 'different' response on an AB or BA trial a *false positive* is a 'same' response on an AA or BB trial.

Separate Analyses of Variance (ANOVAs) were conducted on a) age x ISI x AV/AO modes x clear/noise; b) age x ISI x VO mode x clear/noise; c) ages x ISI x tones contrasts for three modes, and AV-AO augmentation in clear and noise separately. Alpha was set at 0.05, and effect sizes are given for significant differences.

3. RESULTS

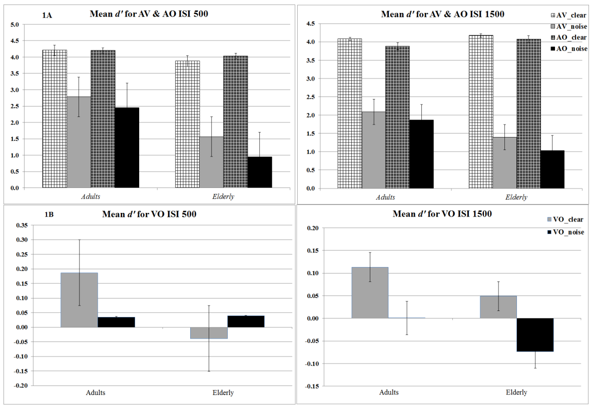
3.1. Overall analyses

Mean *d'* scores over individual tone contrasts for each group are shown for AV/AO and VO scores by ISI and noise condition in Figures 1A and 1B.

3.1.1. Auditory and auditory-visual tone perception: There was a significant main effect of ISI [$F(1,69) = 25.32$, $p<.001$] and an age x ISI interaction [$F(1,69) = 4.88$, $p<.05$]. Figure 1A shows that mean *d'* for adult was higher with ISI 500 (M=3.41) than 1500 (M=2.98), but for the elderly, scores were similar for both ISI (M_{1500}=2.67 & M_{500}=2.60). Also significant were differences between clear (M=4.07) and noise (M=1.77) [$F(1,69) = 442.90$, $p<.001$] and the ISI x noise interaction [$F(1,69) = 21.38$, $p<.001$]. Most importantly there was an AV (M=3.02) vs AO (M=2.81) [$F(1,69) = 8.60$, $p<.01$], and an AV/AO x noise interaction [$F(1,69) = 6.84$, $p<.05$]. Thus there was visual augmentation (AV > AO) evident, especially in the noise conditions, for both adults and the elderly and across both ISIs.

3.1.2. Visual speech (VO) ANOVA: There were no significant effects of any factor on the mean *d'* scores in VO condition. Both groups generally had poor discrimination of tone contrasts when no auditory component was presented, although there appeared to be slightly better performance in the clear condition.

Figure 1: Mean *d'* scores in AV/AO/VO conditions

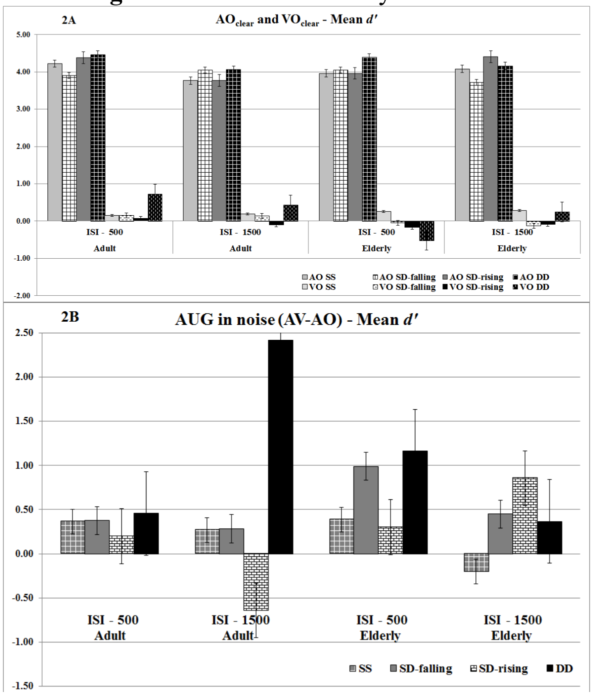


3.2. Relative discriminability of each tone pair

Mean d' scores for 4 types of tone contrasts: SS, SD-falling, SD-rising and DD were examined in more detail, in the AO clear condition; the VO in clear condition; and visual augmentation (AV minus AO) in noise.

3.2.1. AO in Clear condition (see Figure 2A): The FR (DD pair) was significantly more discriminable than all other pairs [$F(1,69) = 14.7$, $p<.001$]; and SD-rising were more discriminable than the SD-falling pairs [$F(1,69) = 8.3$, $p<.01$]. There were significant 2-way interactions of age x SS tone contrasts [$F(1,69) = 4.9$, $p<.05$], and age x SD-rising contrasts [$F(1,69) = 5.2$, $p<.05$] pairs. In all of these cases, adults' scores were higher than those of the elderly. A 3-way age x ISI x the SS pairs interaction [$F(1,69) = 11.1$, $p<.001$] was also significant. These results showed that the predicted SD-rising > SD-falling effect [20] was also evident Thai tone perception across ages. Overall these results can be described as: **DD > SS = SD (with SD-falling < SD-rising)**.

Figure 2: Mean *d'* scores by tone contrasts



3.2.2. VO in Clear condition (see Figure 2A): The only significant VO_{clear} result was a 2-way interaction of age x SD-falling/SD-rising pairs [$F(1,69) = 6.0$, $p<.05$]. In general, both groups found SD-falling slightly more discriminable SD-rising, and this was more pronounced for the adults than the elderly. Thus there was a slight **SD-falling > SD-rising** pattern here.

3.2.3. Augmentation (AV-AO) in Noise (see Figure 2B): The DD pair was slightly more discriminable than the other pairs; and this was more so for ISI 500 than 1500 [$F(1,69) = 5.3$, $p<.05$]. Adults were better at discriminating SD-falling pairs than were the elderly [$F(1,69) = 7.1$, $p<.05$], and the adults also found these pairs more discriminable in ISI 500 while the elderly did better in ISI 1500 [$F(1,69) = 12.7$, $p<.001$]. Thus, there was an overall pattern of **DD > SS = SD (with SD-falling > SD-rising)**.

4. DISCUSSION AND CONCLUSION

This experiment provides strong evidence for visual augmentation (AV > AO) of tone perception, especially in noise, and that this was the case for both Adult and Elderly participants. Results regarding age, visual augmentation, VO perception and particular tone contrasts are set out below.

4.1. Aging

Generally, younger participants were better than the elderly at discriminating lexical tone contrasts in both clear and noise conditions. Discrimination in both clear and noise was better at 500 than 1500 ms ISI for Adults. On the other hand, the Elderly had better discrimination at 1500 than 500 ms in clear speech and equivalent discrimination across ISIs in noise. This indicates that the elderly tend to rely on more phonemic strategies than do young adults. Even though both ages performed worse in noise, the reduction was more pronounced for the elderly.

4.2. Visual augmentation

Visual augmentation of AO tone perception by the addition of visual information (AV > AO) was most noticeable in noise. This was evident across both ages, showing that elderly perceivers, like their younger counterparts, use visual information to enhance speech perception. Thus, visual information is important in tone perception especially in the case of mild to moderate hearing loss and elderly adults are able to use this information effectively.

4.3. VO perception of tones

Neither age group used visual information *alone* (VO) effectively in perception of their native tones. This supports the finding that native Thai listeners as well as listeners of other tone languages perceive VO tone worse than non-tone language listeners [19].

4.4. Tone contrast effects

There were a number of effects specific to the particular tone combinations. Overall the DD (FR pair) was the most discriminable pair compared to the others, while the SS and SD pairs were more or less equally discriminable for the native listeners. Among the SD pairs, SD-rising pairs were better discriminated than SD-falling pairs in the AO_{clear} condition (at ISI 500 for Adults and ISI 1500 for the Elderly, but paradoxically, in VO SD-falling were discriminated better.

In conclusion, the results of this study show that aging results in a deterioration of tone perception, but that elderly Thai perceivers are able to use visual information effectively, just like, but not better than their younger counterparts. One strategy that may be more readily available to elderly perceivers is a more phonemic mode of perception, but further research is required in Thai and in other tone languages before more definitive conclusions can be drawn, or intervention strategies devised.

5. ACKNOWLEDGEMENTS

This research was supported by an Australian Research Council (ARC) Discovery grant (DP0988201) to the last author.

6. REFERENCES

[1] Fromkin, V. 1978. *Tone: A linguistic survey*. New York: Academic Press.

[2] Abramson, A. S. 1978. Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21(4):*319-325.

[3] Henderson, E. J. A. 1981. Tonogenesis: Some recent speculations on the development of tone. *Transactions of the Philological Society, 112:*1-24.

[4] Tseng, C.-Y., Massaro, D. W. and Cohen, M. M. 1985. Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features. *Journal of Chinese Linguistics,* 13:267-289.

[5] Chao, Y.-R. 1930. "A system of tone-letters", *Le Maitre Phonetique*, 45: 24-27.

[6] Werker, J. F., & Logan, J. S. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35–44.

[7] Werker, J. F., & Tees, R. C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.*, 75, 1866-1878.

[8] Campbell, R., Dodd, B., & Burnham, D. (Eds.). 1998. *Hearing by Eye II: Advances in the psychology of speech reading and auditory–visual speech*. East Sussex: Psychology Press.

[9] Vatiktiotis-Bateson, E., Kuratate, T., Munhall, K. G., & Yehia, H. C. 2000. The production and perception of a realistic talking face. In O. Fujimura, B. D. D. Joseph, & B. Palek (Eds.), *LP'98, Item order in language and speech*. Prague: Charles University, Karolinum Press. 439-460.

[10] McGurk, H., & McDonald, J. 1976. Hearing lips and seeing voices. *Nature, 264*, 746–748.

[11] Committee on Hearing, Bioacoustics and Biomechanics (CHABA), 1988. Speech understanding and aging. *J. Acoust. Soc. Am.* 83, 820–895.

[12] Strickland, E.A., Viemeister, N.F., Van Tasell, OJ. & Preminger, J.E. 1994. Is useful speech information carried by fibers with high characteristic frequencies? *J. Acoust. Soc. Am.* 95, 497-501.

[13] Craik, F. I. M., & Masani, P. A. (1967). Age differences in the temporal integration of language. *British Journal of Psychology, 58,* 291-299.

[14] Hummert, M. L., Garstka, T. A., Ryan, E. B. & Bonnesen, J. L. 2004. The role of age stereotypes in interpersonal communication. In J.F. Nussbaum & J. Coupland (Eds.), *Handbook of communication and aging research (2nd ed.,*). Hillsdale, NJ: Lawrence Erlbaum. 91-114.

[15] Burke, D.M. & Shafto, M.A. 2008. *Language and aging*. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition*. New Jersey: Lawrence Erlbaum Associates. 373-443.

[16] Ryan, E.B., See, S.K., Meneer, W.B., & Trovato, D. (1994). Age-based perceptions of conversational skills among younger and older adults. In M.L. Hummert, J.M. Wiemann, & J.N. Nussbaum (Eds.) *Interpersonal communication in older adulthood*. Thousand Oaks, CA: Sage Publications. 15-39.

[17] Baltes, P. B., & Lindenberger, U. 1997. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12,* 12-21.

[18] Wayland, R. P., & Guion, S. G. 2004. Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning*, 54(4). 681-712

[19] Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Xu Rattanasone, N., Schwarz, I.-C., & Webster, D. 2014. Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*. doi:10.1017/S0142716414000496

[20] Krishnan, A., Gandour, J. T., & Bidelman, G. M. 2010. The effects of tone language experience onpitch processing in the brain stem.Journal of Neurolinguistics, 23, 81–95

[21] Sumby, W. H., & Pollack, I. 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am*. 26, 212–215.

[22] Burnham, D., Ciocca, V., & Stokes, S. 2001. Auditory–visual perception of lexical tone. In P. Dalsgaard, B. Lindberg,H. Benner,&Z.-H Tan (Eds.), *Proceedings of the 7th Conference on Speech Communication and Technology, EUROSPEECH 2001 Scandinavia*. 395–398. Retrieved from http://www.isca-speech.org/archive/eurospeech_2001

[23] Mixdorff, H., Hu, Y., & Burnham, D. 2005. Visual cues in Mandarin tone perception. In I. Trancoso, L. Oliviera, & N. Mamede (Eds.), *Proceedings of the 9th European Conference on Speech Communication and Technology*. Bonn, Germany: International Speech. Communication Association. 405-408.

[24] Mixdorff, H., Charnvivit, P., & Burnham, D. K. 2005. Auditory–visual perception of syllabic tones in Thai. In E. Vatikiotis-Bateson, D. Burnham, & S. Fels (Eds.), *Proceedings of the Auditory–Visual Speech Processing International Conference*. Adelaide, Canada: Causal Productions. 3-8.

[25] Smith, D, and Burnham, D. 2012. Facilitation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants. *J. Acoust. Soc. Am.*, 131(2), 1480-1489.

[26] Forster, K. I., & Forster, J. C. 2003. DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers,* 35, 116-124.