Neurobiology of Aging 36 (2015) 283–291

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition

Ryan A. Stevenson[a,b,c,d,*], Caitlin E. Nelms[e,f], Sarah H. Baum[c,g], Lilia Zurkovsky[h], Morgan D. Barense[a,i], Paul A. Newhouse[h], Mark T. Wallace[b,c,d,h,j]

[a] Department of Psychology, University of Toronto, Toronto, Ontario, Canada
[b] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[c] Vanderbilt Brain Institute, Nashville, TN, USA
[d] Vanderbilt Kennedy Center, Nashville, TN, USA
[e] Department of Psychology, Austin Peay State University, Clarksville, TN, USA
[f] Department of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA
[g] Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, TX, USA
[h] Center for Cognitive Medicine, Department of Psychiatry, Vanderbilt University, Nashville, TN, USA
[i] Rotman Research Institute, Toronto, Ontario, Canada
[j] Department of Psychology, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Article history:
Received 20 April 2014
Received in revised form 22 July 2014
Accepted 2 August 2014
Available online 7 August 2014

Keywords:
Speech perception
Multisensory
Aging
Multisensory integration
Inverse effectiveness

ABSTRACT

Over the next 2 decades, a dramatic shift in the demographics of society will take place, with a rapid growth in the population of older adults. One of the most common complaints with healthy aging is a decreased ability to successfully perceive speech, particularly in noisy environments. In such noisy environments, the presence of visual speech cues (i.e., lip movements) provide striking benefits for speech perception and comprehension, but previous research suggests that older adults gain less from such audiovisual integration than their younger peers. To determine at what processing level these behavioral differences arise in healthy-aging populations, we administered a speech-in-noise task to younger and older adults. We compared the perceptual benefits of having speech information available in both the auditory and visual modalities and examined both phoneme and whole-word recognition across varying levels of signal-to-noise ratio. For whole-word recognition, older adults relative to younger adults showed greater multisensory gains at intermediate SNRs but reduced benefit at low SNRs. By contrast, at the phoneme level both younger and older adults showed approximately equivalent increases in multisensory gain as signal-to-noise ratio decreased. Collectively, the results provide important insights into both the similarities and differences in how older and younger adults integrate auditory and visual speech cues in noisy environments and help explain some of the conflicting findings in previous studies of multisensory speech perception in healthy aging. These novel findings suggest that audiovisual processing is intact at more elementary levels of speech perception in healthy-aging populations and that deficits begin to emerge only at the more complex word-recognition level of speech signals.

1. Introduction

Visual cues are known to significantly impact speech perception; when one can both hear a speaker's utterance and concurrently see the articulation of that utterance (lip reading), speech comprehension is more accurate (Ross et al., 2007a, 2011; Sommers et al., 2005; Stevenson and James, 2009; Sumby and Pollack, 1954) and less effortful (Fraser et al., 2010) than when only auditory information is available. The behavioral gain observed when processing information via multiple sensory modalities is governed by a number of factors, with one of the most important being the relative effectiveness of the stimuli that are paired. As a general rule, greater benefits are observed from pairing stimuli that, on their own, are each weakly effective, compared with pairing stimuli that are both strongly effective when presented in isolation (Bishop and Miller, 2009; James et al., 2012; Meredith and Stein, 1983, 1986; Nath and Beauchamp, 2011; Stevenson and James, 2009; Stevenson et al., 2007, 2009, 2012a, 2012b; Wallace et al., 1996; Werner and Noppeney, 2009). This concept of "inverse effectiveness" implies

* Corresponding author at: Department of Psychology, University of Toronto, Rm 523, Sidney Smith, 100 St. George Street, Toronto, ON, Canada M6G 3G3. Tel.: +1 416 978 5464; fax: +1 416 978 4811.
E-mail address: ryan.andrew.stevenson@gmail.com (R.A. Stevenson).

0197-4580/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2014.08.003

that the primary benefits of multisensory integration take place when the individual stimuli provide weak or ambiguous information. For example, the addition of a visual speech signal provides the greatest gain when the auditory speech is noisy (Sumby and Pollack, 1954). Once an individual stimulus is sufficiently salient, the need for multisensory-mediated benefits substantially declines.

In the context of real-world multisensory stimuli, these changes in effectiveness can be mediated not only by changes in the external characteristics of the stimuli (e.g., the loudness of an auditory stimulus) but also by changes in internal events governing the processing of that information. The declines in visual and auditory acuity associated with normal aging are probably the result of decreases in internal signal strength attributable to changes in transduction and encoding processes but also by additional internal noise (i.e., variability) to the transduction and encoding processes. The loss of visual and auditory acuity is seen for both simple and more complex stimuli and is particularly prevalent for speech signals, most notably in the presence of external noise (Dubno et al., 1984; Gosselin and Gagne, 2011; Humes, 1996; Martin and Jerger, 2005; Sommers et al., 2005). Although these age-related declines in speech perception and comprehension have been widely interpreted to be a result of changes in auditory acuity (Liu and Yan, 2007) and diminished ability to filter task-irrelevant auditory information (Hugenschmidt et al., 2009), declines in visual acuity may play an important and underappreciated role. Some evidence suggests that older adults may rely on visual information to a greater extent than their younger counterparts (Freiherr et al., 2013; Laurienti et al., 2006), which may reflect the use of multisensory integration as a compensatory mechanism for declining unisensory abilities.

In this prior work, multisensory gain increased with age for the integration of simple audiovisual stimuli such as flashes and tones, a finding consistent with the principle of inverse effectiveness given the age-related declines in unisensory processing acuity (Laurienti et al., 2006). However, for speech-related stimuli, the picture is more complex and provides only partial support for the concept of inverse effectiveness. Whereas older and younger adults showed equivalent levels of audiovisual gain for high (i.e., easier) signal-to-noise ratio (SNR) trials, older adults showed less gain than younger adults on low (i.e., more difficult) SNR trials (Tye-Murray et al., 2010). One potential explanation for these disparate findings is that in this latter study participants were required to complete or repeat whole sentences, which may introduce variability because of other cognitive factors. For example, verbal memory is known to decline in nondemented aging (Park et al., 2002), a finding that may impair the ability of older participants to recall whole sentences. Thus, the reduced multisensory gain observed on low SNR trials in older adults relative to younger adults may in fact reflect memory impairments for the full sentences, rather than deficits in the integration of auditory and visual cues.

Here, we conducted a novel study designed to address these conflicting observations, specifically structured to examine how aging affects multisensory-mediated gains in speech perception under noisy conditions. Critically, we examined these gains at the level of more elementary (i.e., phonemic) and more complex (i.e., whole word) components of speech, providing the first systematic investigation of how multisensory integration at different levels of processing is affected by aging. We presented younger and older healthy adults with a standard audiovisual speech-in-noise task in which participants reported the perceived word. We used single-word presentations to limit the impact of higher-order cognitive changes known to occur with aging, such as changes in memory or context. Importantly, the task was scored both at the whole-word level and the phoneme level, allowing us to pinpoint whether changes in multisensory gain across the life span differed depending on the level of processing necessary for accurate comprehension. Our results provide evidence that older adults show largely intact multisensory processing at lower (i.e., phonemic) levels of speech perception but begin to show deficits at higher level processing with whole-word recognition.

2. Methods

2.1. Participants

Thirty-four participants (20 females, mean age = 39.0 years, SD = 18.4, range = 19–67 years) completed a behavioral speech-in-noise paradigm. Experimental protocols were approved by Vanderbilt University Medical Center's institutional review board. Participants were divided into 2 age groups based on the median age (39 years) reported in Ross et al. (2007a) publication investigating inverse effectiveness in single-word recognition. The younger group included participants 39 years of age or younger (18 in total, 8 females, mean age = 22.8 years, SD = 4.7, range = 19–38 years), and the older group consisted of participants 40 years of age or older (16 in total, 12 females, mean age = 57.3 years, SD = 6.9, range = 45–67 years). For demographic information, see Table 1. All individuals were screened for normal visual acuity with a tumbling E visual chart and were not hearing-aid users. Additionally, Mini-mental state examinations were conducted on all participants, with a score greater than or equal to 27 used as an exclusionary cutoff, though no participants were excluded. Additionally, participants reported no neurologic impairments. Participants were recruited via flyer at Vanderbilt University Medical Center.

2.2. Stimuli

Stimuli included dynamic, audiovisual (AV) recordings of a female speaker saying 216 tri-phonemic nouns. Stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-talker Database (Sheffert et al., 1996). All stimuli were spoken by speaker F1. The stimuli selected were monosyllabic English words that were matched across sets for accuracy on both visual-only and audio-only recognition (Lachs and Hernandez, 1998) and were also matched across sets in lexical neighborhood density (Luce and Pisoni, 1998; Sheffert et al., 1996). This set of single-word tokens have been used successfully in previous studies of multisensory integration (Stevenson and James, 2009; Stevenson et al., 2009, 2011). Audio signal levels were measured as root mean square contrast and equated across all tokens.

All stimuli throughout the study were presented using MATLAB 2012b (MATHWORKS Inc, Natick, MA, USA) software with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Visual stimuli were 200 × 200 pixels and subtended 10 × 10° of visual angle. Audio stimuli were presented through 2 aligned speakers on each side of the monitor. All tokens lasted 2 seconds and included all prearticulatory gestures.

In the visual-only condition, the visual component of each stimulus or viseme, was presented. Auditory stimuli were all overlaid with 8-channel multitalker babble at 72 dB SPL. The presentation of auditory babble presentation began 500 ms before the beginning of the stimulus token and ended 500 ms following token offset. The root mean square of the auditory babble was linearly ramped up and

Table 1
Participant demographics

	N	Mean age (SD)	Age range (y)	Female (%)	Male (%)
Younger	18	22.8 (4.7)	19–38	44	56
Older	16	57.3 (6.9)	45–67	75	25

Key: SD, standard deviation.