Behav Res (2014) 46:406–415
DOI 10.3758/s13428-013-0372-8

A measure for assessing the effects of audiovisual speech integration

Nicholas Altieri · James T. Townsend · Michael J. Wenger

Published online: 13 August 2013

Abstract We propose a measure of audiovisual speech integration that takes into account accuracy and response times. This measure should prove beneficial for researchers investigating multisensory speech recognition, since it relates to normal-hearing and aging populations. As an example, age-related sensory decline influences both the rate at which one processes information and the ability to utilize cues from different sensory modalities. Our function assesses integration when both auditory and visual information are available, by comparing performance on these audiovisual trials with theoretical predictions for performance under the assumptions of parallel, independent self-terminating processing of single-modality inputs. We provide example data from an audiovisual identification experiment and discuss applications for measuring audiovisual integration skills across the life span.

Keywords Capacity · Integration · Audiovisual gain

A topic of considerable interest in the speech recognition literature concerns how to measure multisensory integration. *Multisensory integration* refers to the ability to benefit from visual speech cues over and above the auditory signal by integrating and effectively combining the two sources of information in support of speech perception. The basic methodology for assessing multisensory integration has typically involved determining whether responses to audiovisual (AV) speech stimuli differ systematically from a function of the responses to the auditory- and visual-only components.

Neuroimaging studies, for instance, have measured blood oxygen level dependent responses to AV speech stimuli in brain regions of interest and have compared them with the maximum unisensory response ($max\{A,V\}$) or the sum of the auditory and visual responses (A + V; e.g., Calvert, Campbell, & Brammer, 2000; Stevenson & James, 2009; Werner & Noppeney, 2010). Likewise, EEG studies have compared peak amplitudes for event-related potentials (ERPs), such as the N100, evoked by auditory and visual stimuli and AV stimuli. Studies have sought to establish evidence for "integration" when the AV peak amplitude systematically differs from the unisensory ERPs (e.g., $AV_{ERP} < A_{ERP} + V_{ERP}$; e.g., Altieri & Wenger, 2013; Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005; Winneke & Phillips, 2011; cf. Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002, for issues concerning anticipatory responses). The logic is that if the multisensory ERP differs statistically from the sum of unisensory ERPs, there is evidence for interactions across modalities and, hence, integration.

Similar logic applies to behavioral measures (accuracy and latency) of AV integration. The question with respect to these measures is whether AV response times (RTs) are shorter or accuracy is greater than on single-modality trials. With respect to the question of whether integration does or does not occur, the question becomes whether AV responses differ from a function of the unisensory responses in such a way as to suggest a violation of independence. A preponderance of studies in the AV speech literature, beginning with Sumby and Pollack's (1954) seminal study, have used mean accuracy to assess the extent one benefits from being able to see a talker's face in noisy listening conditions. Within the domain of RTs, a growing number of studies have used distribution-

N. Altieri (✉)
Department of Communication Sciences and Disorders, Idaho State University, 921 S. 8th Ave. Stop 8116, Pocatello, ID 83209, USA
e-mail: altinich@isu.edu

J. T. Townsend
Department of Psychological and Brain Science, Indiana University, Bloomington, IN, USA

M. J. Wenger
Department of Psychology, The University of Oklahoma, Norman, OK, USA

based measures to assess visual benefit (Altieri & Townsend, 2011; Winneke & Phillips, 2011). However, a unified measure of integration that utilizes both RT and accuracy has rarely been implemented and, to our knowledge, has never been used to assess speech integration. We seek to accomplish that here.[1]

We now introduce a nonparametric measure that assesses performance relative to the predictions of what can be considered a null hypothesis for AV integration: specifically, a model that assumes that the auditory and visual inputs are processed independently and in parallel and that a response can be emitted as soon as either input provides information sufficient for a response (i.e., a parallel independent race model; see Townsend & Nozawa, 1995). Such a model by definition assumes that the auditory and visual modalities do not interact.

Accuracy and response time measures

AV enhancement has often been measured by comparing AV accuracy with auditory-only or, less commonly, visual-only accuracy (e.g., Bergeson & Pisoni, 2004; Erber, 1969; Sumby & Pollack, 1954). The measure of *visual gain*, for instance, has been employed to measure the relative benefit afforded by the visual signal over and above that provided by auditory-only speech information (see Bergeson & Pisoni, 2004; Sumby & Pollack, 1954). Gain is computed by obtaining the mean accuracy in the AV condition and for trials in which only auditory information is available and computing the ratio $\frac{AV-A}{100-A}$. Here, AV denotes percent correct on AV trials, A represents the percent correct on auditory-only trials, and V is percent correct on visual-only trials. The predicted AV accuracy, assuming independence between auditory and visual modalities (i.e., without integration occurring), is given by the probability: $p(\mathrm{AV})=p(\mathrm{A})+p(\mathrm{V}) - p(\mathrm{A}) * p(\mathrm{V})$. In this equality, $p(\mathrm{A})$ denotes the observed probability correct for auditory-only trials, $p(\mathrm{V})$ for visual-only trials, and $p(\mathrm{AV})$ denotes the predicted probability correct on AV trials. The accuracy and gain measures have been utilized in several studies assessing speech integration in normal hearing, as well as aging and clinical populations with either sensory or cognitive deficits (Bergeson & Pisoni, 2004; Erber, 2002; Grant, Walden, & Seitz, 1998; Massaro, 2004; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sommers, Tye-Murray, & Spehar, 2005).

With respect to measures based on RT, integration has been assessed by computing a measure from the redundant-targets literature known as the race model, or Miller's inequality. This inequality is constructed using the empirical cumulative distribution functions (CDFs) for RTs in each of the different classes of trials: $F_{\mathrm{A}}(t)+F_{\mathrm{V}}(t)-F_{\mathrm{AV}}(t)\geq 0$, where $F_{\mathrm{A}}(t)$ and $F_{\mathrm{V}}(t)$ are the CDFs for the auditory-only and visual-only trials, respectively, and $F_{\mathrm{AV}}(t)$ is the CDF for the AV trials (e.g., Altieri & Townsend, 2011; Besle et al., 2004; Winneke & Phillips, 2011). Violations of this inequality (i.e., values <0) may be interpreted as evidence of facilitative auditory–visual interactions, a form of parallel processing known as *coactivation* (for discussions, see, e.g., Colonius & Townsend, 1997; Miller, 1982; Townsend & Nozawa, 1995), or integration of the auditory and visual inputs.

Altieri and Townsend (2011) extended the RT-based approach by characterizing integration in terms of the construct of *capacity* (e.g., Townsend & Ashby, 1978; Townsend & Nozawa, 1995). The capacity coefficient allows one to compare processing times from trials where both auditory and visual information are availablewith RTs obtained from auditory- and visual-only trials using parallel race model predictions as a benchmark.[2] This characterization of capacity is based on the distributions of RTs, rather than a single moment of the distribution (e.g., the mean) and, as such, provides increased sensitivity and precision (see relevant discussions in, e.g., Townsend & Ashby, 1978, 1983; Townsend & Wenger, 2004; Wenger & Gibson, 2004; Wenger, Negash, Petersen, & Petersen, 2010). Capacity is constructed by obtaining the empirical CDF of the RTs [$F(t)$] in each condition and transforming it into the integrated hazard function by this identity: $H(t)=-\log[1 - -F(t)]$. The latter identity facilitates estimation of the latent underlying integrated hazard function, since the term, $-\log[1-F(t)]$, is straightforward to estimate (see, e.g., Altieri & Townsend, 2011; Townsend & Nozawa, 1995).

This function can be interpreted in terms of the cumulative amount of work done or energy expended by time t in each of the three trial conditions (Townsend & Ashby, 1978,

[1] Interestingly, procedures have been introduced in the RT literature to correct for "fast guesses" that tend to occur at the tail end of the RT distribution (see, e.g., Eriksen, 1988; Gondan & Heckel, 2008; Miller & Lopes, 1991; see also Pachella, 1974, for a combined treatment of RTs and accuracy). A certain proportion of fast guesses is assumed to be correct simply by chance and may introduce a bias that inflates redundancy gain due to overestimation of the CDF (Eriksen, 1988). Some suggested correction methods include obtaining RT estimates from incorrect responses and subtracting it from a proportion of correct responses, or alternatively, trimming the tail end of the RT distributions (Rach et al., 2010). These procedures for correcting CDFs may be used to estimate the hazard functions shown in the capacity equations in the following section (see Eq. 1). Still, this methodology differs from the logic introduced in the following section, in which obtained accuracy is included directly in the integration assessment function.

[2] This class of parallel models assumes that, on average, the speeds of individual channels are not affected by whether or not other channels are in operation—thus, *unlimited capacity*. The classical independent race model is a special case of unlimited capacity channels, which further assumes stochastic independence across channels.

1983; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Thus, with this index of cumulative work, it is possible to compare the total amount of work done on the AV trials at each point in time with the combined cumulative work done on the auditory-only and visual-only trials using a *capacity coefficient* (Altieri & Townsend, 2011; Townsend & Nozawa, 1995; Townsend & Wenger, 2004):

$$C(t) = \frac{H_{AV}(t)}{H_A(t) + H_V(t)}. \quad (1)$$

Since $C(t)$ is a ratio, three possible outcomes emerge (Townsend & Eidels, 2011; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). First, capacity can be greater than 1 at a specific point in time, indicating faster responses on AV trials, as compared with race model predictions derived from auditory- and visual-only trials. If this occurs, integration is said to be efficient, suggesting *supercapacity*, because responses in the AV condition are faster than independent race model predictions. Second, $C(t)$ can be less than 1. This indicates inefficient or *limited capacity* integration due to longer RTs in the AV condition, as compared with predictions from the A-only and V-only conditions. Third, $C(t)$ can be equal to 1 at time t. This would indicate that AV processing is just as efficient as processing auditory or visual inputs alone and would be interpreted as *unlimited capacity* integration.[3]

A critical strength of the capacity coefficient is that it can be formally related to other distribution-based measures relative to capacity (see Townsend & Eidels, 2011, for a unifying discussion). In particular, the values of $C(t)$ can be related to known upper and lower limits, under assumptions of independence, such that values of $C(t)$ that exceed these bounds guarantee departures from unlimited capacity processing. Deviations from unlimited capacity processing (under the assumption of parallel independent channels) could be produced by inhibitory or excitatory cross-modal interactions (which we would predict for multisensory integration). Such interactions could be due to temporal correlations between processing times, such as might be the result of attentional allocation strategies in difficult listening conditions. For example, the participant may concentrate on the A modality when useful and on the V modality when useful, thereby inducing negative channel correlation causing $C(t)>1$ (see, e.g., Colonius, 1990; Colonius & Vorberg, 1994; Mordkoff & Yantis, 1991). Efficient allocation of attention could produce supercapacity, since the listener "knows" when to tune into the visual modality for complementary cues, and also when to ignore it. Selectively attending to only one modality—say, the auditory—while completely ignoring the visual would normally cause $C(t)<1$. This is because AV trials would yield similar RTs to the auditory-only condition. Finally, deviations could also exist because of different processing architectures, such as serial or coactive (Townsend & Nozawa, 1995; Townsend & Wenger, 2004).

The capacity coefficient presented in Eq. 1 assumes that the RTs used for estimation are all observed on trials on which a correct response was generated. This is a standard assumption in the treatment of RT data. However, by definition, it excludes information about response accuracy. For both theoretical and applied reasons, it would be beneficial to have a characterization that takes into account both accuracy and latency information (e.g., Ratcliff, Thapar, & McKoon, 2004) and the ability to obtain information from the signal appear to be adversely affected by the aging process and/or diminished sensory acuity.

Combining response time and accuracy

In this section, we introduce the (capacity) *integration assessment measure*, $C\_I(t)$, which takes into account both RTs and accuracy; like the RT-only measure, $C\_I(t)$ is distribution-free and nonparametric. We show that incorporating both speed and accuracy into a capacity measure is a nontrivial exercise requiring modified $F(t)$ distribution functions.

The combined use of RT and accuracy is based on a foundational logic similar to that used in Eq. 1, because performance on AV trials is compared with independent race model predictions. The details and underlying theory for the RT and accuracy approach can be found in Townsend and Altieri (2012), and the logic is as follows. Suppose a listener is presented with an AV stimulus. The listener can correctly identify the word if he or she correctly identifies the auditory, visual, or both auditory and visual information by a certain time (e.g., Altieri & Townsend, 2011). The probability that they correctly identify the target word in the auditory or visual input by some time t can occur through the following sum of likelihoods, shown in Eq. 2: being correct on A at or before time t while being incorrect on V, being correct on V at or before time t while being incorrect on A, being correct on A at or before time t and correct on V (while recognition has not yet occurred on V), being correct on V at or before time t and correct on A (while recognition has not yet occurred on A), and being correct and complete on both A and V by time t.

[3] The observed $F(t)$s are assumed to include a residual motor component. One way in which this residual component has been dealt with in the literature is to assume the existence of a base time distribution that is convolved with processing time. Potential effects that base processing time may have on observed capacity have been discussed by Townsend and Honey (2007).

$$\int_0^t P_{AV}(T_{AVC}=t'\cap t'<T_{AVI})dt' = \int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*P_V(I)+\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt'*P_A(I)$$
$$+\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*\int_{t'=t}^{\infty} P_V(T_{VC}=t'\cap t'<T_{VI})dt'$$
$$+\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt'*\int_{t'=t}^{\infty} P_A(T_{AC}=t'\cap t'<T_{AI})dt'$$
$$+\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt' \quad (2)$$

Eq. 2 appears complex, but it is actually very manageable. Each term is approximately identical to a cumulative distribution (frequency) function, $F(t)$, that can be obtained from RT data from A-only, V-only, and AV trials. Consider the following term from Eq. 2: $\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'$. This denotes the empirical CDF from A-only trials on which the participant responds correctly (since the correct process reaches threshold before the incorrect one does), weighted by the overall probability of being correct on those trials. Here, the integral indicates a cumulative value from time 0 to time t. The result is similar to a CDF, but instead of summing to 1, it integrates to the probability correct on A-only trials. Specifically, $P_A(\text{`` ''})$ in the integral denotes the probability of being correct by a certain time t on A-only trials (in which the time to process the correct word, T_{AC}, is faster than the time to process the incorrect word, T_{AI}).

Consider another term $\int_{t'=t}^{\infty} P_V(T_{VC}=t'\cap t'<T_{VI})dt'$. This term is equal to the probability that the listener will make a correct V-only identification of the stimulus by time t weighted by the probability correct on visual-only trials but that a correct response has not been made by time t. This term is similar to $1-F(t)$ (the probability that processing has not finished by a certain time), except that it begins at the overall probability correct on visual-only trials instead of 1. $P_V(\text{`` ''})$ denotes the probability of being correct on V-only trials (again, the probability that the word actually presented, $T_{VC}$, is recognized in the visual domain before an incorrect word, $T_{VI}$). Finally, the same is true for $P_{AV}(\text{`` ''})$, which represents the probability of being correct on AV trials. Lastly, the terms $P_A(I)$ and $P_V(I)$ denote the probability of being incorrect on auditory- and visual-only trials, respectively. Similar to the original capacity coefficient, obtaining estimates of these terms from empirical data is relatively straightforward (see the Appendix).

Computing the integration coefficient (Eq. 3) involves obtaining the logarithm of each term in Eq. 2 and then dividing the independent model prediction derived from unisensory trials in the numerator by the observed AV data, $\int_0^t P_{AV}(T_{AVC}=t'\cap t'<T_{AVI})dt'$, in the denominator. Faster and more accurate responses on AV trials, as compared with A- and V-only trials, yield values larger than 1, which implies increasingly efficient and accurate integration. If the accuracy weightings are all allowed to go to 1, then Eq. 2 becomes Eq. 3, the original capacity function showing that $C(t)$ is a special case of the more general integration assessment function; specifically, $C(t)$ constitutes a special case of $C\_I(t)$ in which no errors are made and incorrect categories are not accounted for in the race.

$$C\_I(t)=\frac{\log\left[\begin{array}{l}\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*P_V(I)+\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt'*P_A(I)\\ +\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*\int_{t'=t}^{\infty} P_V(T_{VC}=t'\cap t'<T_{VI})dt'+\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt'*\int_{t'=t}^{\infty} P_A(T_{AC}=t'\cap t'<T_{AI})dt'\\ +\int_0^t P_A(T_{AC}=t'\cap t'<T_{AI})dt'*\int_0^t P_V(T_{VC}=t'\cap t'<T_{VI})dt'\end{array}\right]}{\log\left[\int_0^t P_{AV}(T_{AVC}=t'\cap t'<T_{AVI})dt'\right]} \quad (3)$$

For purposes of illustration, Table 1 shows a truth table corresponding to the accuracy values assuming an OR decision rule. In the next section, we illustrate the use of the integration assessment function (Eq. 3) with data from one example participant in an AV speech identification study that measured integration efficiency under variable listening conditions (Altieri, 2011).

Example data: Application across different sensory Conditions

The $C\_I(t)$ measure provides information about whether deviations were observed from independent model predictions at each point of the RT distribution (i.e., for both fast and slow responses). Critically, these deviations could be the result of either faster or slower AV processing speed, higher or lower AV accuracy, or a combination of RT and accuracy. For diagnostic purposes, we suggest the strategy of also computing $C(t)$ to determine whether AV RT influenced redundant target performance, and also computing accuracy scores (auditory, visual, and AV) to examine whether differences between predicted and obtained AV accuracy may have influenced performance. We provide MATLAB code in the Appendix to illustrate the computation of these measures.

The example task required participants to make eight-alternative forced choice identification responses to monosyllabic spoken words, by way of a buttonpress response on a keyboard (Altieri, 2011; see also Altieri & Townsend, 2011, for methods). The numeric keys 1–8 (top row of a standard keyboard) were labeled with high-frequency monosyllabic words, and participants became acquainted with their arrangement during a practice session at the beginning of each block. The set of words included "boat," "date," "gain," "mouse," "page," "job," "shop," and "tile." The videos were of two female talkers from the Hoosier Multi-Talker Database. Accuracy and RTs were obtained in AV, auditory (A), and visual-only (V) blocks.[4] Listeners were presented with three auditory S/N ratios (clear, −12 dB, and −18 dB SPL mixed with white noise) to investigate the effects that variable listening conditions have on AV integration. A total of 240 trials were obtained in each condition in separate blocks (AV_{Clear}, AV_{-12}, AV_{-18}, A_{Clear}, A_{-12}, A_{-18}, and V-only blocks).

Table 1 The left column ("Auditory") indicates whether the stimulus word (i.e., "Correct") is recognized first in the race or whether an incorrect is (i.e., "Incorrect"). The second column shows the same for the visual modality. The third column, labeled "Winner," indicates whether recognition first occurs in the auditory or visual modality. Finally, the fourth column, labeled "Accuracy," indicates the accuracy of the response given the information in the other columns

Auditory	Visual	Winner	Accuracy
Correct	Correct	Auditory	Correct
Correct	Correct	Visual	Correct
Correct	Incorrect	Auditory	Correct
Correct	Incorrect	Visual	Incorrect
Incorrect	Correct	Auditory	Incorrect
Incorrect	Correct	Visual	Correct
Incorrect	Incorrect	Aud./Vis.	Incorrect

The $C(t)$ and $C\_I(t)$ results are shown for the example participant in Fig. 1. The results for the RT-only measure of capacity (Eq. 1) for each auditory S/N ratio are shown in the left panel. The results for $C\_I(t)$ (Eq. 3) are shown in the right panel. The presentation of both $C(t)$ and accuracy scores allows one to separately analyze the contribution of RT and accuracy on efficiency. Each point indicates the $C(t)$ and $C\_I(t)$ value across each time point for three auditory S/N ratios. The results suggest that efficiency, in terms of speed and accuracy, improved in the AV condition, as compared with the unisensory conditions, as the auditory signal became increasingly degraded. In the clear auditory condition, the results showed, not surprisingly, that the visual information failed to have any facilitatory effect either on speed or on accuracy. This is evidenced by the fact that $C(t)$ and $C\_I(t)$ were considerably less than 1 for a large range of RTs.

Table 2 shows the accuracy scores from each condition and each auditory S/N ratio (A and AV), as well as the race model predictions (A+V−AV). Visual-only accuracy was 69 % correct. When the auditory S/N ratio was clear, the obtained AV accuracy was nearly identical to race model predictions. However, in the −12- and −18-dB conditions, obtained AV scores were greater than race model predictions for accuracy: $p(AV)=p(A)+p(V) - p(A) * p(V)$.

The RT results (Eq. 1) in the left panel indicate inefficient integration when the auditory S/N ratio was clear [$C(t)<1$]. This is due to the fact that the listener failed to benefit from visual information under optimal listening conditions. The integration coefficient in the right panel (Eq. 3) was similar to $C(t)$ [i.e., $C\_I(t)<1$]. Since the obtained AV accuracy approximated race model predictions, the decrement in efficiency observed in both panels results from a slowdown, relative to independent race model predictions.

[4] One reason a block design was implemented, rather than a mixed design, was to maximize ecological validity. The number of modalities available does not typically change suddenly in everyday conversation. For example, listeners must regularly communicate in environments that are auditory-only, such as the telephone. We also point out that previous AV speech research has compared early neural measures of integration (e.g., N1/P2) along with behavioral measures in a block versus mixed design and has not obtained significant differences (van Wassenhove et al., 2005; see Bergeson & Pisoni, 2004; Sommers et al., 2005; and Winneke & Phillips, 2011, for other examples of block designs in AV speech tasks). Finally, since the experimental trials are presented in blocks, probabilistic contingencies (Mordkoff &Yantis, 1991), which could facilitate or inhibit capacity levels, were likely avoided.

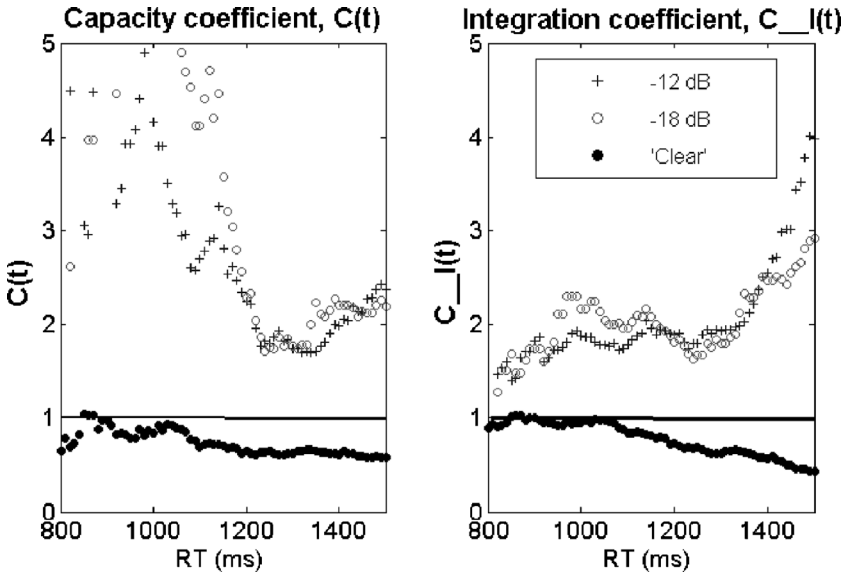

Fig. 1 Traditional capacity coefficient (left) and integration assessment measures (right) from a listener in a word identification task. Data are shown across three different auditory S/N ratios, which can be used to simulate sensory deficit. The symbols indicate the function values for a specific auditory S/N ratio at a certain time point

For the lower auditory S/N ratios, processing speed regularly violated race model predictions [$C(t)>1$]. In fact, as auditory listening conditions deteriorated from −12 to −18 dB, integration efficiency increased. A qualitatively similar pattern of results can be observed in the right panel of Fig. 1, where the integration coefficient, $C\_I(t)$, was measured [i.e., $C\_I(t)>1$]. Because AV accuracy violated race model predictions and the RT measure of $C(t)$ was much greater than 1, it shows that AV benefit under degraded listening conditions is manifested through enhancement of both speed and accuracy. Nonetheless, $C\_I(t)$ results differ from $C(t)$ in important ways. For the shorter RTs, $C(t)$ indicates supercapacity (>5 for the low S/N ratios). While $C\_I(t)$ also suggests supercapacity for this same time range, estimated capacity is lower (~2). Thus, $C(t)$ shows that while the AV RTs were much faster than model predictions, the obtained AV accuracy moderated integration efficiency as estimated by taking accuracy into account using $C\_I(t)$. These data show a young, normal-hearing listener with efficient integration in the RT domain but less efficient integration measured by $C\_I(t)$ due to suboptimal gain in accuracy (Table 2). Contrast this with an aging or hearing-impaired listener with poor auditory- and visual-only recognition who may show substantially higher gain in the accuracy domain (e.g., Bergeson & Pisoni, 2004). Now, if such a listener were to take advantage of AV processing speed similar to the listener in Fig. 1 (see Winneke & Phillips, 2011), then $C\_I(t)$ might show greater supercapacity than $C(t)$. On the other hand, if the listener slows down substantially on AV trials to achieve greater multisensory accuracy, $C\_I(t)$ may take a hit. Overall, the new $C\_I(t)$ measure can elucidate valuable quantitative information through the addition of accuracy information that RT alone simply cannot.

Table 2 Auditory-only and audiovisual (AV) accuracy scores for each condition, with race model predictions for accuracy

S/N Ratio	A-Only	Obtained AV	Race Model
Clear	.98	.98	.99
−12 dB	.69	.95*	.90
−18 dB	.36	.90*	.80

Note. The "*" denotes an obtained accuracy level greater than race model predictions.

Applications to sensory decline

In this section, we discuss predictions for integration efficiency for listeners experiencing sensory decline. We focus particularly on predictions related to speed and accuracy for aging populations, although similar predictions may be derived for other clinical populations, such as individuals with cochlear implants.

Age-related decline in sensory and cognitive function, as well as individual differences in sensory and cognitive ability, can affect how listeners perceive information from auditory and visual speech signals (e.g., Bergeson & Pisoni, 2004; Erber, 2002; Sommers et al., 2005; Winneke & Phillips, 2011). The aging process adversely affects both auditory and visual functioning. This is evidenced by poorer lip-reading skills in aging participants, as compared with their younger counterparts (Sommers et al., 2005; Winneke & Phillips, 2011), and by difficulty in auditory-only speech perception as a result of age-related hearing loss (e.g., Erber, 2002, 2003). Although the effects of age-related sensory declines in each sensory modality are well established, the influences of sensory functioning on

one's ability to integrate auditory and visual speech signals is at best ambiguous.

One reason for this ambiguity is that research has yielded divergent conclusions regarding integration ability. This is partially due to the fact that different studies have utilized different dependent measures, with some studies relying only on accuracy (e.g., Sommers et al., 2005) and others relying on RTs (Winneke & Phillips, 2011). Some studies that have compared AV integration and visual gain (in terms of accuracy) in young normal-hearing listeners with those in elderly adults have failed to show evidence for superior multisensory speech recognition in aging populations. Sommers et al. (2005) obtained accuracy scores in older and younger adults and adjusted the auditory S/N ratio individually to control for auditory recognition performance across individuals and age groups. The participants subsequently participated in recognition tasks using syllables, words, and sentences involving auditory-only, visual-only, and AV trials. The authors observed slightly poorer AV enhancement in aging populations, possibly because older adults exhibited poorer visual-only recognition skills, as compared with younger normal-hearing individuals. Thus, older adults may benefit less from visual speech information, as compared with younger listeners.

Conversely, some evidence suggests that the ability to benefit from visual information may be superior in older listeners, as compared with their younger normal-hearing counterparts (e.g., Bergeson & Pisoni, 2004; Laurienti, Burdette, Maldjian, & Wallace, 2006). In a review of the effect of aging on AV speech integration, Bergeson and Pisoni provided evidence showing that although unisensory function declines with age, the ability to benefit from visual speech cues (e.g., place of articulation; Grant et al., 1998) and combine them with degraded auditory information actually improves. AV enhancement might be greater in older adults, particularly when auditory recognition levels are not controlled across age groups. As is shown in the example data, the ability to integrate auditory and visual speech cues becomes especially important under difficult sensory conditions. This is in line with the law of *inverse effectiveness*: As auditory-only and visual-only recognition become less effective, AV recognition improves relative to unisensory recognition, likely due in part to the reduction of ceiling effects (e.g., Altieri & Townsend, 2011; Laurienti et al., 2006; Ross et al., 2007; Stein & Meredith, 1993).

Interestingly, a recent study using converging measures of AV processing (RTs and ERPs) reported similar benefit in the RT distributions across age groups. The data revealed violations of Miller's (e.g., 1982) upper bound on processing efficiency and race model predictions (Winneke & Phillips, 2011). The violation of that bound implies that the AV benefit, in terms of speed, was greater than predicted by a substantial class of parallel models. On the other hand, their analysis of the N1/P1 ERP components did show important differences across age groups. Both groups showed evidence for an AV amplitude reduction. That is, the value of the AV peak was less than that of the A + V only peak, suggesting interactions across modalities. However, the degree of the reduction in the early sensory P1 component was significantly larger in older adults. The authors argued that the findings were consistent with the hypothesis that fewer neural resources were required to yield equivalent behavioral performance in older adults.

Winneke and Phillips's (2011) findings point to a contradiction between neural and behavioral data. On the one hand, the ERP data suggest more efficient use of neural resources in older than in younger adults. On the other hand, the behavioral data indicate identical or very similar AV enhancement across age groups. In their study, the auditory S/N ratio was elevated for elderly participants in order to equate auditory accuracy levels with their younger counterparts (~85 %). This essentially meant that more auditory cues were available, on average, for elderly participants. The higher auditory clarity may have altered early sensory coding in older participants, causing changes in very early ERP. Thus, predictive visual information might not have facilitated low-level early sensory encoding due to the increased availability of auditory cues. One possible explanation for the lack of evidence for improvement in the behavioral data is that elderly participants might have lowered their decision criteria on AV trials because the relative increase in availability of low- and high-frequency auditory cues helped them benefit from visual information (e.g., Erber, 2003; although see Ratcliff et al., 2004). This may have helped them achieve a similar level of performance to younger listeners in the RT domain.

The implementation of the $C\_I(t)$ and $C(t)$ measures should, in future studies, separate out the effects of speed and accuracy on one's ability to combine visual speech with auditory information pertaining to manner, nasality, and vowels. Suppose we equate auditory S/N ratio across age groups in a paradigm similar to the detection design described in the previous section. Elderly or hearing-impaired listeners should generally process unisensory auditory and visual information less accurately, as compared with younger listeners (e.g., Erber, 2003; Sommers et al., 2005). However, if $C\_I(t)$ shows similar efficiency across age groups, two testable predictions emerge:

1. Listeners may utilize more conservative decision criteria on AV trials (in this case, slowing RTs) in order to achieve a greater accuracy gain to equate performance. Hence, these listeners may slow down on AV trials in order to benefit from complementary cues; $p(AV)$ should exceed race model predictions, although $C(t)$ may show limitations.
2. Combined AV information facilitates processing in the time domain. However, this would be done at the expense of accuracy. The $C(t)$ analysis should indicate greater or similar gain, as compared with younger listeners, although the accuracy data should show less gain.

Given that RT data fit to diffusion models (e.g., Thapar et al., 2003) have consistently pointed to more conservative

decision strategy in elderly participants, we believe that the first prediction is certainly possible for some listeners (although cf. Winneke & Phillips, 2011). However, *C_I*(*t*) bears implications and is a potential testing ground for all parameterized models, in addition to Ratcliff's diffusion model.

Finally, *C_I*(*t*) may differ for aging versus younger participants depending on their history. A lower *C_I*(*t*) could be the result of the inability to benefit in terms of speed and/or accuracy from visual speech cues. This situation may arise in very low auditory S/N ratios and, especially, in older adults with a high degree of hearing impairment because of the inability to obtain low-frequency auditory speech information (Erber, 2003). Conversely, *C_I*(*t*) may reveal superior integration in certain listeners if they benefit from the visual signal in terms of speed (e.g., Laurienti et al., 2006), compensate in the accuracy domain, or both. We predict that this situation may arise in some older adults with progressive hearing loss and better lipreading ability who have extensive practice associating lip movements with auditory vowel and consonant cues in order to maintain efficient face-to-face communication skills.

Summary and conclusion

Recent studies have demonstrated a promising approach for evaluating integration efficiency in populations with deficits in sensory functioning (Altieri, 2011; Altieri & Townsend, 2011; Winneke & Phillips, 2011). The approach of assessing integration efficiency can be significantly improved upon by utilizing our nonparametric measure of integration efficiency—*C_I*(*t*)—one that takes into account both accuracy and RT in a single measure. We provided example data from a closed set speech detection experiment to illustrate its utility. This study simulated sensory decline by introducing changes in auditory S/N ratio, and as was predicted, *C_I*(*t*) [and *C*(*t*)] indicated significant enhancement in integration efficiency in terms of speed and accuracy under diminished sensory functioning (e.g., Ross et al., 2007; Stein & Meredith, 1993). Future studies obtaining normative data may combine this measure with open set sentence processing measures (e.g., Sommers et al., 2005) in order to obtain a converging estimate of integration efficiency.

Taken together, the new integration assessment coefficient, *C_I*(*t*), can be used to capture both individual and group differences in the ability to integrate multisensory speech information simultaneously in terms of their accuracy plus RTs. This approach will provide the speech research community with a yardstick for assessing integration efficiency by taking into account time-based and accuracy measures relative to parallel race model predictions.

Consider, for example, how some research indicates that aging populations integrate speech information less efficiently than do younger normal-hearing listeners (e.g., Sommers et al., 2005). The combined *C_I*(*t*), RT, and accuracy approach introduced here can provide a more comprehensive diagnosis the locus of such effects: (1) Does sensory decline associated with aging contribute to a decline in sensitivity, mainly causing accuracy to take a hit, or (2) might there be higher level changes (e.g., decision criteria) that contribute to changes in AV processing rate as well? How does variability in listening conditions covary with changes in integration efficiency across age groups? Finally, the integration assessment function may reveal connections between brain signals and behavior in both normal-hearing and aging populations.

Author Note This research was supported by funds provided to the Visual Neuroscience Laboratory at The University of Oklahoma, by the Dept. of Psychology, and by the INBRE Program at Idaho State University (NIH Grant Nos. P20 RR016454 (National Center for Research Resources) and P20 GM103408 (National Institute of General Medical Sciences).

Appendix

```
% Example Matlab Code for C(t) and C_I(t)
% Clear Auditory Condition
clc; format short g;
t=0:10:3000; % Time vector in msec (MIN : bin size : MAX)
num=240; % number of trials per each condition
% ################## read data ####################
% Read in Data Vector: 720 trials
% Input data in Vector format
% ################################################
% dividing the trials into three conditions based on
%Presence/Absence
%Input Accuracy Levels for A, V, and AV:
A=.98; V=.69; AV=.98; % Example from Clear Auditory Cond.
rt1=data((1):num); % Auditory Only: 240 Trials
rt2=data((num+1):2*num); % Visual Only: 240 Trials
rt3=data((2*num+1):3*num); % Audiovisual: 240 Trials
% trim the data
rt1trim=rt1(rt1>100 & rt1<3000);
rt2trim=rt2(rt2>100 & rt2<3000);
rt3trim=rt3(rt3>100 & rt3<3000);
% estimating the functions
% nnz(rt#) gives the number of nonzero elements
f1=hist (rt1trim, t) / nnz(rt1trim); % Density
F1=cumsum (f1); % Cum. frequency
f2=hist (rt2trim, t) / nnz(rt2trim);
F2=cumsum (f2);
f3=hist (rt3trim, t) / nnz(rt3trim);
F3=cumsum (f3);
% Calculating capacity
C=log(1-F3) ./ (log(1-F2)+log(1-F1)); % C(t)
% Find the Capacity measures for correct Responses
F1NC=cumsum(f1)*A;
```

```
F2NC=cumsum(f2)*V;
F3NC=cumsum(f3)*AV;
F1NCS=(1-cumsum(f1))*(A);
F2NCS=(1-cumsum(f2))*(V);
F3NCS=(1-cumsum(f3))*(AV);
CORII=log(F1NC*(1-V)+F2NC*(1-A)+
F1NC.*F2NCS+F2NC.*F1NCS+F1NC.*F2NC)./log(F3NC);
subplot(1,2,1)
hold on
scatter(t, C, 15, 'filled')
title2 ('Capacity coefficient, C(t)','FontSize',14,'FontWeight',
'bold')
xlabel ('RT (ms)','FontSize',14); ylabel ('C(t)','FontSize',16);
axis ([100 2500 0 5])
subplot(1,2,2)
hold on
scatter(t, CORII, 15, 'filled')
title ('Assessment Coefficient, A(t)','FontSize',14,'FontWeight',
'bold')
xlabel ('RT (ms)','FontSize',14); ylabel ('C(t)','FontSize',16);
axis ([100 2500 0 5])
```

References

Altieri, N. (2011). Neural and information processing measures of audiovisual integration. *Journal of Vision, 11*(11), 791.

Altieri, N. & Wenger, M. (under review). Neural dynamics of audiovisual integration efficiency under variable listening conditions.

Altieri, N., & Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Frontiers in Psychology, 2*(238), 1–15.

Bergeson, T. R., & Pisoni, D. B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 153–176). Cambridge, MA: The MIT Press.

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20,* 2225–2234.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649-657.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34,* 253–275.

Colonius, H. & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice*, *Decision and Measurement*, volume in honor of R. Duncan Luce, Mahwah, NJ: Erlbaum Associates.

Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology, 38,* 35–58.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12,* 423–425.

Erber, N. P. (2002). *Hearing, vision, communication and older people*. Clifton Hill, Victoria, Austrailia: Clavis Publishing.

Erber, N. P. (2003). Use of hearing aids by older people: influence of non-auditory factors (vision, manual dexterity). *International Journal of Audiology, 42*(2), S21–S25.

Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics, 44*(2), 191–193.

Gondon, M., & Heckel, A. (2008). Testing the race model inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology, 52*(5), 322–325.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*(5), 2677–2690.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*(8), 1155–1163.

Massaro, D. W. (2004). From multisensory integration to talking heads and language learning. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 153-176). Cambridge, MA: The MIT Press.

Miller, J. O., & Lopes, A. (1991). Bias produced by fast guessing in distribution-based tests of race models. *Perception & Psychophysics, 50,* 584–590.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 520–538.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing, 41-82*. Potomac, MD: Lawrence Erlbaum.

Rach, S., Diederich, A., Steenken, R., & Colonius, H. (2010). The race model inequality for censored reaction time distributions. *Attention Perception & Psychophysics*, *72*(3), 839-847.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50,* 408–424.

Ross, L. A., Saint-Amour, D., Leavitt, V., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I'm saying? Optimal Visual Enhancement of Speech Comprehension in noisy environments. *Cerebral Cortex, 17*(5), 1147–53.

Sommers, M., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear & Hearing, 26,* 263–275.

Stein, B. E., & Meredith, M. A. (1993). *Merging of the senses*. Cambridge, MA: MIT Press.

Stevenson, R.A., & James, T.W. (2009). Neuronal convergence and inverse effectiveness with audiovisual integration of speech and tools in human superior temporal sulcus: Evidence from BOLD fMRI. *NeuroImage*, *44*, 1210-1223.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 12–15.

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14,* 106–114.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, *18*, 415-429.

Townsend, J. T., & Altieri, N. (2012). A capacity assessment function that measures performance relative to standard parallel predictions. *Psychological Review, 119*(3), 500–516.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan and F. Restle (Eds.), *Cognitive Theory Vol. III* (pp. 200-239). Hillsdale, NJ: Erlbaum Associates.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge: Cambridge University Press.

Townsend, J.T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin & Review*, *18*, 659-681.

Townsend, J. T., & Honey, C. J. (2007). Consequences of base time for redundant signals experiments. *Journal of Mathematical Psychology, 51*(4), 242–265.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology, 39,* 321–360.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11,* 391–418.

van Wassenhove, V., Grant, K., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Science, U.S.A., 102,* 1181–1186.

Wenger, M. J., & Gibson, B.S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 708-719.

Wenger, M. J., Negash, S., Petersen, R. C., & Petersen, L. (2010). Modeling and estimating recall processing capacity: Sensitivity and diagnostic utility in application to mild cognitive impairment. *Journal of Mathematical Psychology, 54,* 73–89.

Werner, S. & Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, *20*(8), 1829-1842.

Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging, 26,* 427–438.