Integration efficiency for speech perception within and across sensory modalities by normal-hearing and hearing-impaired individuals

Ken W. Grant[a)]
Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, D.C. 20307-5001

Jennifer B. Tufts
Department of Communication Sciences, University of Connecticut, Storrs, Connecticut 06269

Steven Greenberg
Silicon Speech, 46 Oxford Drive, Santa Venetia, California 94903

(Received 22 October 2006; revised 9 November 2006; accepted 15 November 2006)

In face-to-face speech communication, the listener extracts and integrates information from the acoustic and optic speech signals. Integration occurs within the auditory modality (i.e., across the acoustic frequency spectrum) and across sensory modalities (i.e., across the acoustic and optic signals). The difficulties experienced by some hearing-impaired listeners in understanding speech could be attributed to losses in the extraction of speech information, the integration of speech cues, or both. The present study evaluated the ability of normal-hearing and hearing-impaired listeners to integrate speech information within and across sensory modalities in order to determine the degree to which integration efficiency may be a factor in the performance of hearing-impaired listeners. Auditory-visual nonsense syllables consisting of eighteen medial consonants surrounded by the vowel [a] were processed into four nonoverlapping acoustic filter bands between 300 and 6000 Hz. A variety of one, two, three, and four filter-band combinations were presented for identification in auditory-only and auditory-visual conditions: A visual-only condition was also included. Integration efficiency was evaluated using a model of optimal integration. Results showed that normal-hearing and hearing-impaired listeners integrated information across the auditory and visual sensory modalities with a high degree of efficiency, independent of differences in auditory capabilities. However, across-frequency integration for auditory-only input was less efficient for hearing-impaired listeners. These individuals exhibited particular difficulty extracting information from the highest frequency band (4762–6000 Hz) when speech information was presented concurrently in the next lower-frequency band (1890–2381 Hz). Results suggest that integration of speech information within the auditory modality, but not across auditory and visual modalities, affects speech understanding in hearing-impaired listeners. [DOI: 10.1121/1.2405859]

PACS number(s): 43.71.An, 43.71.Ky, 43.71.Es [ADP] Pages: 1164–1176

I. INTRODUCTION

The ability to understand speech relies to a considerable degree on the integration of spectro-temporal information from different regions of the acoustic frequency spectrum. This cross-spectral integration is especially important for individuals using multichannel hearing aids and cochlear implants. Such auditory prostheses partition the spectrum into separate channels and subject them to various forms of signal processing. However, because most spoken conversations involve face-to-face interaction, visual speech information is often available in addition to auditory information. In these cases, listeners integrate information from both the auditory and visual modalities. This cross-modal integration occurs in tandem with the cross-spectral integration occurring entirely within the auditory modality. The present study evaluated the efficiency of cross-modal and cross-spectral integration of speech information in normal-hearing and hearing-impaired listeners. This was done by analyzing error patterns in consonant identification tests consisting of spectrally sparse stimuli presented in auditory-only and auditory-visual conditions. A visual-only condition was also included.

When speech is presented acoustically, it is spectrally filtered in the auditory pathway into many overlapping frequency channels. Information about specific speech sounds may be distributed broadly across many different frequency channels or may be concentrated within a limited number of channels. Information distributed across many frequency channels may be helpful for decoding the speech signal, particularly in noisy and reverberant conditions. Early work on speech vocoder systems (Hill *et al.*, 1968) showed that speech recognition improved as the number of discrete frequency channels increased from three to eight (for frequencies between 180 and 4200 Hz). More recently, Shannon *et al.* (1995) demonstrated that the temporal information contained in as few as four broad spectral regions is sufficient

[a)]Author to whom correspondence should be addressed. 301 Hamilton Avenue, Silver Spring, Maryland 20901. Telephone: 202-782-8596. Electronic mail: grant@tidalwave.net

for good speech recognition in ideal listening conditions. These studies were conducted in quiet using normal-hearing subjects whose abilities to integrate information across spectral channels were not in doubt. However, it is unclear whether individuals with sensorineural hearing loss are able to integrate information across spectral channels as efficiently as normal-hearing listeners do, especially if the hearing loss varies widely across the speech frequency range. Turner *et al.* (1999) addressed this issue, using speech signals composed primarily of temporal cues from different spectral bands. They assessed the intelligibility of 1, 2, 4, and 8-channel conditions. They hypothesized that normal-hearing and hearing-impaired listeners would perform similarly when speech was limited to a small number of spectral channels. As more spectral channels were used to represent the speech information, and the frequency resolution of the hearing-impaired listeners became a limiting factor, intelligibility for hearing-impaired listeners would fall below that of normal-hearing listeners. The results showed that normal-hearing and hearing-impaired listeners performed similarly when only one spectral channel was used, as expected. However, contrary to prediction, the hearing-impaired listeners were significantly poorer than normal-hearing listeners at identifying the speech tokens for all other conditions, including the two-band condition where frequency resolution would not be expected to be a problem. These results are consistent with those of Friesen *et al.* (2001) comparing speech recognition performance for normal-hearing subjects and cochlear implant (CI) patients as a function of the number of spectral channels. Friesen *et al.* (2001) demonstrated that most CI subjects were unable to utilize fully the spectral information provided by the electrodes used in their implants.

As Turner *et al.* (1999) point out, these results are difficult to interpret. Hearing-impaired and normal-hearing listeners have seemingly comparable abilities to extract temporal cues from the speech envelope derived from a single auditory channel (though single-channel speech decoding is poor for both groups of listeners). However, when even minimal amounts of spectral information are introduced into the speech signal (as when temporal envelopes from two different spectral regions are combined), hearing-impaired listeners perform more poorly than normal-hearing listeners. Although hearing-impaired listeners usually have poorer-than-normal spectral resolution, there are no models of hearing impairment that limit listeners with moderate sensorineural hearing loss to a single auditory channel. Thus, it is unclear why, once audibility has been taken into account, hearing-impaired listeners fail to benefit from the addition of a second channel of speech information to the same degree as normal-hearing listeners.

One explanation offered by Turner *et al.* (1999) was that greater overlap between adjacent bands may have existed due to broader critical bands in hearing-impaired listeners. This overlap can be likened to a third channel that introduces noise or unusable temporal cues. To test this hypothesis, Turner *et al.* (1999) interposed a band-reject region between 1000 and 2000 Hz that separated the two spectral bands. In this way, energetic interference between the two spectral bands would be greatly reduced. It was anticipated that the speech recognition performance of the hearing-impaired listeners would improve in this condition. However, performance in both conditions (with and without spectrally contiguous channels) was essentially the same. A second explanation for the failure of hearing-impaired listeners to take advantage of additional spectral channels of information posited the existence of some form of central auditory deficit in these listeners due to their greater age (45–70 years) compared with the normal-hearing subjects (22–48 years).

The question remains as to why hearing-impaired listeners fail to benefit from the addition of a second channel of speech information to the same degree as normal-hearing listeners. It is possible that, although the hearing-impaired listeners were able to extract similarly useful information from the individual channels, they were not able to integrate the information across these channels as efficiently as the normal-hearing listeners. Healy and colleagues (Healy and Bacon, 2002; Healy, Kannabiran and Bacon, 2005) have presented additional evidence that hearing-impaired listeners have difficulty integrating temporal speech information presented in different frequency regions. They tested word recognition of normal-hearing and hearing-impaired listeners using two sinusoidal signals (750 and 3000 Hz) modulated by one-third octave bands of speech and presented either synchronously or asynchronously. Word recognition was essentially zero for either channel presented alone. The frequencies of the sinusoids were chosen to maximize channel isolation, and the presentation levels were chosen to ensure audibility. Listeners with hearing impairment performed more poorly than listeners with normal hearing, even at comparable sensation levels, suggesting that a deficit existed in their ability to integrate temporal speech information across different frequency regions. Furthermore, when between-band asynchrony was introduced, the performance of the hearing-impaired listeners fell far more precipitously than that of the normal-hearing listeners, and more sharply than predicted based on correlations of the envelopes of the two bands (Healy, Kannabiran and Bacon, 2005). These results provide further evidence for a deficit in cross-spectral integration separate from other effects of sensorineural hearing loss such as reduced audibility and broadened auditory filters.

Combining speech cues derived from multiple sources of information has been a topic of considerable discussion and research, primarily with respect to integration of cues across the auditory and visual sensory modalities (Massaro, 1987, 1998; Braida, 1991; Grant and Seitz, 1998; Massaro and Cohen, 2000; Grant, 2002). Auditory-visual integration refers to the process of combining information that has been extracted from the auditory and visual channels (Grant, 2002). All other things being equal, greater skill at integrating auditory and visual cues, or higher integration efficiency, will almost always lead to better performance in auditory-visual tasks (Grant, Walden, and Seitz, 1998). Highly efficient integrators are assumed to be better at using cues from multiple sources for speech recognition.

Models of auditory-visual integration conceptualize the extraction and integration of cues as independent processes

that operate serially. The models also assume no interference across modalities in the extraction of cues. However, listeners may not function this way in the real world. Sommers *et al.* (2005a) presented evidence that auditory-visual integration, as assessed with the Prelabeling Model of Integration model (Braida, 1991; see Sec. II for further information on this model), varied depending on the signal-to-noise ratio (SNR) at which the speech materials were presented. Sommers *et al.* (2005a) interpreted these results to mean that the processes of extraction and integration of unimodal cues interacted to determine overall auditory-visual benefit. Their conclusion appears related to that of Ross *et al.* (2006), who assessed unimodal and auditory-visual word recognition over a wide range of SNRs. They argued that auditory-visual integration is most effective for intermediate SNRs. That is, a minimum amount of auditory information is required before word recognition can be most effectively enhanced by visual cues. In extremely favorable or unfavorable SNRs, in which one modality is clearly dominant over the other, listeners may rely less heavily on the integration of cues across modalities to understand speech. Instead, they may switch to a strategy in which a strong bias exists for cues from the dominant modality, and information from the less-dominant modality is discarded. This form of "interference" in cue extraction across modalities would not be accounted for by current models and would lead to the appearance of sub-optimal auditory-visual integration. Nevertheless, models of auditory-visual integration assess cue extraction and integration independently. This is necessary because, in the absence of a comprehensive understanding of the interactions that may occur across such processes, integration cannot be validly assessed without first determining which cues are available for integration.

Models that were originally developed for assessing auditory-visual integration efficiency can also be applied to cross-spectral integration in the auditory pathway (Greenberg *et al.*, 1998; Silipo *et al.*, 1999; Müsch and Buus, 2001; Greenberg and Arai, 2004). In either the cross-modal or cross-spectral case, predictions of an optimal processing model are compared to actual performance, and the differences are used as an index of integration efficiency. For example, if observed performance is just slightly worse than predicted performance, then integration efficiency is considered nearly, but not quite, optimal. By examining the differences between observed and predicted performance, it may be possible to shed light on why listeners with sensorineural hearing loss do not benefit as much as expected by the addition of information in other spectral channels.

In the present study, the integration efficiency of normal-hearing and hearing-impaired listeners was compared both within and across modality. Nonsense syllable (VCV) tokens were spectrally filtered into four non-overlapping bands. These bands were combined in various ways and then presented to normal-hearing and hearing-impaired subjects for identification. Auditory-only and auditory-visual conditions were examined. The spectral band configurations have been used in earlier, studies of spectro-temporal integration (Greenberg *et al.*, 1998; Silipo *et al.*, 1999; Greenberg and Arai, 2004) and were chosen so as to preclude ceiling effects when several bands were combined or when visual cues were available in addition to information in one or more of the bands.

Previous research (Grant and Walden, 1996; Grant *et al.*, 1998) has shown that speech-reading conveys significant information about consonantal place of articulation (i.e., [b] versus [d]), relatively little information about manner of articulation (e.g., [b] versus [m], especially once place cues have been accounted for), and virtually no information about voicing (e.g., [p] versus [b]). Acoustically, place information is usually associated with second and third format transitions (Pickett, 1999) in the range between 1000 and 2500 Hz. Thus, the information conveyed by visual cues is essentially redundant with that provided by the acoustic signal in the mid-frequency region. If mid-frequency acoustic information is removed, but visual cues are available, efficient cross-modal integration should result in good speech recognition performance. Conversely, if only mid-frequency acoustic information is available along with visual cues, even the most efficient cross-modal integration would be expected to produce only modest gains in performance. The auditory-visual conditions of the experiment were designed with these expectations in mind. The purpose of the study was to address the following question: Are decrements in speech recognition performance in hearing-impaired listeners attributable to a deficit solely in the extraction of speech cues, or is the decrement due to a deficit in integration efficiency within and/or across sensory modalities, or both?

II. METHODS

A. Subjects

Four normal-hearing (mean age=43 years, range =29–55 years) and four hearing-impaired subjects (mean age=71 years, range=65–74 years) were recruited from the staff and patient population of the Army Audiology and Speech Center, Walter Reed Army Medical Center. Pure tone thresholds for right and left ears at frequencies between 250 and 8000 Hz are shown in Table I. For normal-hearing subjects, thresholds were 20 dB HL or better at all test frequencies, with the exception of one subject who had a threshold of 25 dB HL at 8000 Hz (ANSI, 1989). For hearing-impaired subjects, the mean three-frequency threshold at 500, 1000, and 2000 Hz was 39.2 and 33.3 dB HL for right and left ears, respectively. The average high-frequency threshold at 3000, 4000, and 6000 Hz was 72.1 and 75 dB HL for right and left ears, respectively. Recent audiometric evaluations showed immittance measures within normal limits and no significant air-bone gaps, indicating that the hearing loss was sensorineural in origin. The hearing-impaired subjects were experienced hearing-aid users who had received audiological examinations and hearing-aid fittings at the Army Audiology and Speech Center, Walter Reed Army Medical Center. All subjects were native speakers of American English with normal or corrected-to-normal vision (visual acuity equal to or better than 20/30 as measured with a Snellen chart). Subjects were compensated for their participation, as permitted by federal regulations. Each subject read and signed an informed consent form prior to beginning the

TABLE I. Pure tone thresholds (dB HL) for four normal-hearing subjects (NH) and four hearing-impaired subjects (HI).

	RIGHT EAR									LEFT EAR								
	250	500	1000	1500	2000	3000	4000	6000	8000	250	500	1000	1500	2000	3000	4000	6000	8000
NH1	5	10	5	15	15	15	0	5	25	5	10	0	10	10	15	5	5	5
NH2	15	15	10	10	15	20	15	15	20	10	5	5	10	10	15	10	20	15
NH3	5	0	5	5	0	5	10	5	20	10	5	5	5	0	0	10	10	10
NH4	5	5	5	5	15	15	5	10	10	5	0	0	5	20	10	5	0	15
HI1	20	20	20	15	35	55	65	70	70	20	15	15	35	40	95	95	100	90
HI2	20	25	40	45	55	70	80	80	80	20	20	25	40	35	60	65	80	80
HI3	10	25	50	50	45	50	60	105	120	20	15	45	45	40	55	60	75	75
HI4	40	40	45	70	70	75	75	80	75	30	35	45	65	70	70	70	75	70

study. The methods used in this study were approved by the Institutional Review Board at the Walter Reed Army Medical Center.

B. Stimuli

Consonant recognition in a vowel-consonant-vowel context was evaluated using spectrally sparse acoustic stimuli consisting of one, two, three, or four narrow (1/3-octave) spectral bands separated by an octave (e.g., the upper edge of one band was an octave below the lower edge of its higher-frequency neighbor). The stimuli consisted of the consonants /b,p,g,k,d,t,m,n,f,v,θ,ð,s,z,ʃ,ʒ,tʃ,dʒ/ surrounded by the vowel /ɑ/. Ten unique productions of each nonsense syllable were spoken by a female speaker of American English and recorded audiovisually using a three-tube Ikegami color camera and stored on optical disk (Panasonic TQ-3031F). Two of these ten tokens were selected for training and the remaining eight tokens were reserved for testing. The speech tokens were then processed through a MATLAB© routine to create filtered speech tokens containing one to four spectrally distinct 1/3-octave bands. FIR filters were used with attenuation rates that were, at minimum, 100 dB/octave (and usually between 500 dB/octave and 2000 dB/octave). The four-band auditory condition ($A_{1,2,3,4}$) consisted of filter passbands of 298–375 Hz (band 1), 750–945 Hz (band 2), 1890–2381 Hz (band 3), and 4762–6000 Hz (band 4) presented simultaneously. Four additional auditory conditions were tested, which included band 1 alone (A_1), bands 1 and 4 combined ($A_{1,4}$), bands 2 and 3 combined ($A_{2,3}$), and bands 1, 2, and 3 combined ($A_{1,2,3}$). Two separate auditory-visual conditions were tested in which subjects viewed a video image of the talker presented synchronously with either the two fringe bands ($AV_{1,4}$) or the two middle bands ($AV_{2,3}$). An eighth condition (V) examining visual-only speech recognition (i.e., speechreading) was also tested. The experimental conditions are listed in Table II.

The audio portion of each production was digitized at a sampling rate of 20 kHz with 16-bit amplitude resolution. The digitized samples were ramped on and off (using a 50 ms raised cosine function), lowpass-filtered at 8.5 kHz, and normalized in level so that all stimuli had the same average rms amplitude. One effect of this normalization was to alter the levels of the bands in the sub-band conditions relative to the $A_{1,2,3,4}$ condition. A sixth auditory condition ($A_{1,2,3,4sum}$) was added later to evaluate the effect of these band-level differences. This condition is the sum of the $A_{1,4}$ and $A_{2,3}$ conditions, and was not normalized in level. Thus, the band levels in this condition are equal to the band levels in the $A_{1,4}$ and $A_{2,3}$ conditions, but are more intense than the band levels in the $A_{1,2,3,4}$ condition. Table III shows the relative band levels for each of the auditory conditions. For auditory-visual presentations, the digitized computer audio

TABLE II. Experimental conditions. Performance was predicted for the conditions marked in parenthesis using Braida's Prelabeling Model of Integration. Predicted and obtained results were then compared to derive Integration Efficiency measures (i.e., Obtained /Predicted ×100). See text for further explanation.

	Condition	Description
1	($A_{1,2,3,4}$)	Consonants filtered into four non-overlapping narrow filter bands: Band 1 (298–375 Hz), Band 2 (750–945 Hz), Band 3 (1890–2381 Hz), and Band 4 (4762–6000 Hz).
2	$A_{1,4}$	Consonants filtered into two bands: Bands 1 and Band 4.
3	$A_{2,3}$	Consonants filtered into two bands: Bands 2 and Band 3.
4	A_1	Consonants filtered into one band: Band 1.
5	$A_{1,2,3}$	Consonants filtered into three bands: Band 1, Band 2, and Band 3.
6	$A_{1,2,3,4sum}$	Additive sum of Condition 2 and Condition 3. This condition was more intense than the $A_{1,2,3,4}$ condition, especially for Bands 1 and 4. See text for further explanation.
7	($AV_{1,4}$)	Condition 2 presented auditory-visually.
8	($AV_{2,3}$)	Condition 3 presented auditory-visually.
9	V	Consonants were presented visually only (i.e., speechreading).

TABLE III. Relative band levels for each of the four filtered-speech bands in the six auditory conditions.

	RELATIVE BAND LEVELS (dB)			
CONDITION	BAND 1	BAND 2	BAND 3	BAND 4
A_1	−8.04			
$A_{1,4}$	−9.42			−22.22
$A_{2,3}$		−9.13	−27.26	
$A_{1,2,3}$	−17.05	−10.52	−28.71	
$A_{1,2,3,4}$	−17.46	−10.59	−29.55	−30.76
$A_{1,2,3,4sum}$	−9.45	−10.15	−28.40	−22.30

and optical disk video portions of each production were realigned using custom auditory-visual control software. Alignments were verified using a dual-trace oscilloscope to compare the original and digitized productions of each utterance and were found to be accurate within ±2 ms. Video signals from the optical disk were routed through a digital time-base corrector (FOR. A FA-310) before being sent to a 21″ color monitor (SONY PVM 2030) situated approximately 1.5 m from the subject. When active, the video monitor displayed a life size image of the talker's face, neck, and shoulders.

C. Procedure

Subjects were seated in a double-walled sound-attenuating booth. A touch-screen terminal was placed within easy reach of the subject and displayed the full set of 18 consonants used. All audio test signals were output through a 16-bit DAC (TDT DD1) and routed through an 8.5 kHz anti-aliasing filter (TDT FLT3), separate programmable attenuators (TDT PA4), mixer (TDT ADD1), head phone driver (TDT HBUF3), and stereo headphones (Beyer Dynamic DT770). Speech signals were presented binaurally at approximately 85 dB SPL for normal-hearing subjects and at a comfortable level (between 95 and 105 dB SPL depending on the subject) for hearing-impaired subjects. A third-octave band analysis was conducted on the speech stimuli for condition $A_{1,2,3,4}$ calibrated to 85, 95, and 105 dB SPL in order to determine the audibility of the bands. At these presentation levels, all of the speech information in bands 1, 2, and 3 would have been audible to the hearing-impaired subjects (with the exception of the weaker portions of the signal in band 3 for subject ECC). Given the severity of the hearing losses in the higher frequencies, it is likely that band 4 was not fully audible in some cases. Specifically, the mean rms SPL in band 4 (approximately 71–81 dB, depending on presentation level) was close to the thresholds of some of the hearing-impaired subjects; thus, while the speech peaks in band 4 would have been audible, the weaker portions of the speech signal may have been inaudible. Each trial began by first informing the subject of the test modality (auditory, visual, or auditory-visual) and then by playing a warning tone followed by one of the nonsense syllables (drawn from the set of eight unique productions of each syllable). Subjects responded by pressing one of the 18 consonants displayed on the touch screen, and then pressed a second touch area marked with the word "continue" when they were ready for the next trial. Subjects could change their responses as often as they wished until the continue button was pressed. Each block consisted of 72 trials of a single condition (four repetitions of each consonant × 18 consonants). Each condition with the exception of $A_{1,2,3,4sum}$was tested ten times, yielding a total of 40 trials (four repetitions × 10 blocks) per consonant per condition. The order of auditory, visual, and auditory-visual conditions was randomized for each subject. No feedback was provided. Subject responses were stored in the form of stimulus-response confusion matrices for subsequent analysis. Testing and analysis of the control condition $A_{1,2,3,4sum}$ was completed following the other eight conditions.

TABLE IV. Feature classification for voicing, manner, and place categories.

VOICING	
Voiced:	b,d,g,m,n,v,ð,z,ʒ,dʒ
Unvoiced:	p,t,k,f,θ,s,ʃ,tʃ
MANNER OF ARTICULATION	
Stop:	b,p,g,k,d,t
Nasal:	m,n
Fricative:	v,f,ð,θ,z,s,ʒ,ʃ
Affricate:	dʒ,tʃ
PLACE OF ARTICULATION	
Bilabial:	b,p,m
Alveolar:	d,t,n,s,z
Labio-Dental:	v,f
Dental:	ð,θ
Palatal:	ʒ,ʃ,dʒ,tʃ
Velar:	g,k

D. Data analysis and model fits

For each of the conditions listed in Table II, excluding $A_{1,2,3,4sum}$, performance measures were computed for overall consonant recognition, and information transmission (Miller and Nicely, 1955) for the articulatory-acoustic features of voicing, manner of articulation, and place of articulation. Information-transmission rates for the phonetic features listed in Table IV were obtained using the SINFA algorithm (Wang, 1976), with values taken from the summary table for unconditional feature information.

A common metric for evaluating auditory-visual benefit, and sometimes auditory-visual integration (Sommers *et al.*, 2005b), is the relative benefit measure (RB) described by Sumby and Pollack (1954) This metric evaluates the improvement in information transmission that occurs when auditory and visual information is available, relative to the case when only auditory or only visual information is available. Given an auditory-visual condition and some reference condition (either auditory-only or visual-only), the RB is the observed percent improvement in the amount of information transmitted for the auditory-visual condition compared with the information transmitted for the reference condition, divided by the theoretically possible percent improvement [e.g., $(AV-A)/(100-A)$], where AV is the percent information transmitted in the auditory-visual condition and A is the percent information transmitted in the auditory-only reference condition. Maximum RB is one. The RB is described here because of its historical importance as a measure of auditory-visual benefit, and because it offers an opportunity to examine why large auditory-visual benefit does not necessarily imply good auditory-visual integration, and vice versa. Consider a hypothetical auditory-only condition in which only place-of-articulation information is conveyed to the listener and a score of A ($A \ll 100\%$) is obtained. Next, consider the addition of visual cues to the auditory-only condi-

tion. Since speech-reading conveys primarily place cues, the auditory and visual information will be highly redundant. Little to no information about voicing or manner will be conveyed in the auditory-visual condition. A score of 100% information transmission is not possible in this case (Grant *et al.*, 1998). Therefore, the theoretically possible improvement is not 100-*A*, but some smaller number *Y*-*A*, $Y<100$. If the listener perfectly integrates all of the available auditory and visual information, he or she will only obtain a score of *Y* and a corresponding RB of (*Y*-*A*)/(100-*A*). Note that the denominator in the RB measure is 100-*A*, not Y-*A*, yielding an RB less than one. If RB is used as a measure of *AV* integration, this listener will appear to have less-than-optimal integration ability, when in actuality he or she has performed as optimally as possible in that condition.

Furthermore, the use of whole-word percent correct scores, as in the RB measure, does not provide enough information to calculate the theoretically possible improvement. A finer analysis of error patterns in terms of the phonetic features of voicing, manner and place of articulation is required.

Models for assessing integration efficiency have been developed that do not share these limitations (Massaro, 1987; Blarney *et al.*, 1989; Braida, 1991). In the present study, integration efficiency was assessed with the prelabeling (PRE) model of integration (Braida, 1991). In this model, the information the listener has extracted from the signal and which is available for integration is recovered by analyzing the error patterns generated in unimodal (i.e., auditory-only and visual-only) conditions. Next, this information is combined using an "optimal" decision rule that assumes perfect integration of the available information, and that also assumes an unbiased receiver with no interference across auditory frequency channels or across auditory and visual modalities. The resulting AV score is a prediction of optimal performance in the auditory-visual condition, given the information extracted from the auditory-only and visual-only channels. By comparing predicted and obtained scores, a metric for integration efficiency is computed. The PRE model subjects the confusion matrices from the unimodal conditions to a special form of multidimensional scaling (MDS) that is interpreted within a theory of signal detection (e.g., Green and Swets, 1966; Macmillan *et al.*, 1988). The model provides a spatial interpretation of the ability to distinguish between consonants, analogous to that derived from traditional MDS (Borg and Lingoes, 1987). However, unlike traditional MDS, the scaled distances between consonants in the separate condition spaces are converted to a common metric, d', which explicitly reflects the correctness of response (and thus compensates for potential response bias). The decision rule assumes a comparison between stimulus attributes (modeled as a multidimensional vector of cues, $\vec{X}$) and prototypes or response centers ($\vec{R}$) in memory. Subjects are assumed to respond R_k if the distance from the observed vector of cues $\vec{X}$ to $\vec{R}_k$ is smaller than the distance to any other prototype. A subject's sensitivity $d'(i,j)$ in distinguishing stimulus S_i from stimulus S_j is given by

$$d'(i,j) = \|\vec{S}_i - \vec{S}_j\| = \sqrt{\sum_{k=1}^{D} (S_{ik} - S_{jk})^2}, \quad (1)$$

where $\|\vec{S}_i - \vec{S}_j\|$ is the distance between the *D*-dimensional vector of cues generated by stimuli S_i and S_j.

In the present study, estimates of stimulus and response centers that best fit a given confusion matrix were obtained iteratively using a KYST procedure (Kruskal and Wish, 1978). For the first iteration, $\vec{S}$ and $\vec{R}$ are assumed to be aligned. Subsequent iterations attempted to improve the match between predicted and obtained matrices (using a χ^2 measure) by displacing slightly both stimulus and response centers. Each iteration assumed 5120 presentations per consonant token yielding a total of 92, 160 trials per matrix (i.e., 18 consonants $\times$ 5120 presentations). This number was selected to reduce the stimulus variability in each MDS fit to approximately 1/10th of the variability in the data. The MDS fits were further optimized by choosing either two- or three-dimensional solutions depending on which gave the best fit to the unimodal matrix.

PRE model predictions for $A_{1,2,3,4}$, $AV_{1,4}$, and $AV_{2,3}$ performance were made solely on the basis of performance in the $A_{1,4}$, $A_{2,3}$, and *V* conditions. Assuming that speech cues from different frequency bands and from the visual signal are combined optimally, the decision space for the combined conditions is the Cartesian product of the space for each of the component conditions. Thus, the relation between a subject's sensitivity in an auditory-visual condition (e.g., $AV_{1,4}$) and the corresponding unimodal sensitivities (e.g., $A_{1,4}$ and *V*), assuming no perceptual interference (e.g., masking or distraction) across modalities, is given by

$$d_{AV_{1,4}}(i,j) = \sqrt{d_{A_{1,4}}(i,j)^2 + d_V(i,j)^2}. \quad (2)$$

Similar equations may be written to describe model predictions for the $AV_{2,3}$ and $A_{1,2,3,4}$ conditions.

Predictions for the $A_{1,2,3,4}$, $AV_{1,4}$, and $AV_{2,3}$ conditions were compared to actual performance exhibited by individual subjects. Since the PRE model is an optimum-processor model, predicted scores should always equal or exceed observed scores. A subject's integration efficiency, as calculated by the model, is given by the ratio between observed and predicted recognition scores expressed as a percentage (with 100% indicating perfect integration).

III. RESULTS

Consonant recognition scores for the eight different conditions listed in Table II (excluding $A_{1,2,3,4\text{sum}}$) are illustrated in Fig. 1. The top panel shows the mean data for the five auditory-only conditions whereas the bottom panel shows the mean data for the visual-only and auditory-visual conditions. Not surprisingly, auditory-only recognition performance for the hearing-impaired subjects was worse than that of the normal-hearing subjects for all conditions except for A_1. The band in the A_1 condition was in a spectral region (298–375 Hz) where threshold differences between the normal-hearing and hearing-impaired groups were relatively small. In spite of the reduced recognition scores of the hearing-

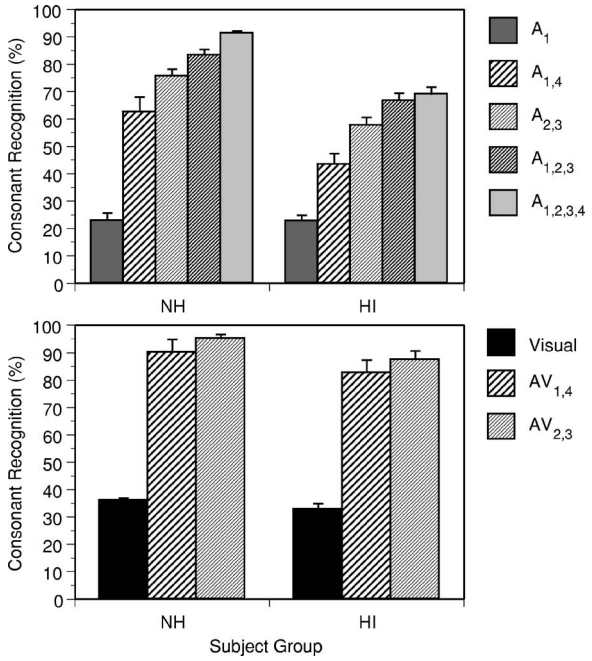

FIG. 1. Average consonant recognition scores for the five auditory conditions (top) and the three visual and auditory-visual conditions (bottom). Data are for four normal-hearing (NH) subjects and four hearing-impaired (HI) subjects. Error bars show +1 standard error.

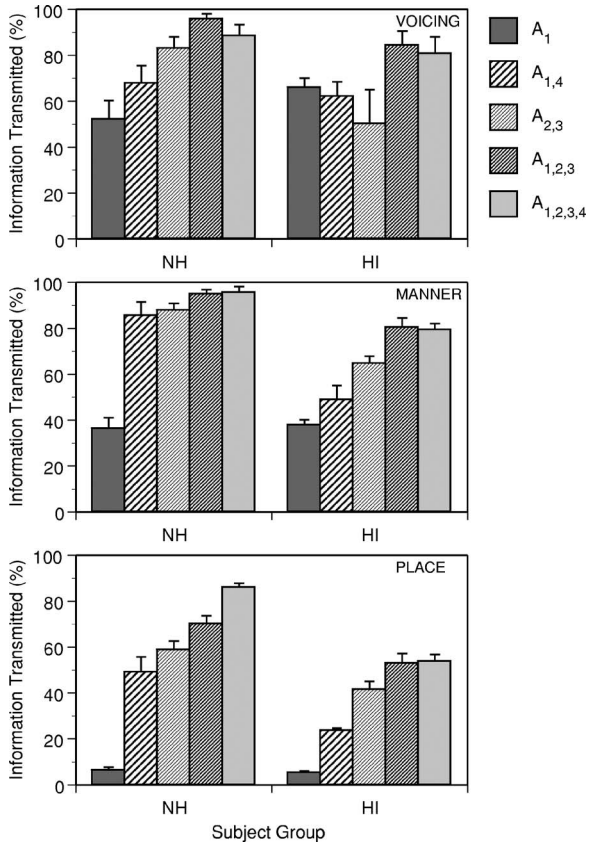

FIG. 2. Percent information transmitted in the five auditory conditions for the speech features voicing (top), manner-of-articulation (middle), and place-of-articulation (bottom). Data are for four normal-hearing (NH) subjects and four hearing-impaired (HI) subjects. Error bars are +1 standard error.

impaired subjects in the other auditory-only conditions, the overall pattern of performance between the two groups was generally similar. However, hearing-impaired subjects showed proportionally less benefit than normal-hearing subjects when band 4 was added to either band 1 alone (i.e., $A_{1,4}$ compared to A_1) or to bands 1, 2, and 3 combined (i.e., $A_{1,2,3,4}$ compared to $A_{1,2,3}$). In particular, the amount of benefit afforded by band 4 in the context of the first three bands (i.e., $A_{1,2,3,4}$ versus $A_{1,2,3}$) was negligible for the hearing-impaired listeners. The bottom panel of Fig. 1, on the other hand, shows that visual-only and auditory-visual recognition performance was comparable across subject groups, suggesting that most of the hearing deficit was overcome when speechreading was combined with even limited auditory information.

The data displayed in Fig. 1 were subjected to a repeated-measures ANOVA with group (normal-hearing or hearing-impaired) as a between-subjects factor and condition as a within-subjects factor. Two separate analyses were conducted, one for the auditory-only conditions and one for the visual-only and auditory-visual conditions. For the auditory-only conditions, the main factors of group and condition, as well as their interaction, were all significant [group: $F(1,6)=19.48, p=0.005$; condition: $F(4,24)=250, p<0.001$; group*condition: $F(4,24)=8.99, p<0.001$]. The significant interaction arises from the already noted difference in the ability of listeners to make use of the information in band 4, with hearing-impaired listeners deriving less advantage than normal-hearing listeners. For the three conditions involving speechreading (V, $AV_{1,4}$, and $AV_{2,3}$), only the effect of condition was significant [$F(1,6)=444, p<0.001$], this being driven by the large differences between the visual-only and the two auditory-visual conditions. *Post hoc* tests indicated that there were no significant differences between the groups for either auditory-visual condition, or between the separate auditory-visual conditions within groups. Thus, in spite of significant differences in auditory-only recognition performance across the two subject groups (as seen in the upper panel of Fig. 1), once the visual signal was combined with the acoustic speech signal, all subjects recognized the consonants with nearly equal accuracy (as seen in the lower panel of Fig. 1). For example, although the hearing-impaired subjects' mean score in condition $A_{1,4}$ was significantly lower than the mean score for the normal-hearing subjects, their mean score in condition $AV_{1,4}$ (in which visual cues were added to $A_{1,4}$) was not significantly different from the normal-hearing subjects' mean score. This finding also held for conditions $A_{2,3}$ and $AV_{2,3}$. This apparent increase in auditory-visual benefit for the hearing-impaired subjects relative to the normal-hearing subjects (i.e., equivalent auditory-visual and visual-only performance across groups but reduced auditory-only performance in the hearing-impaired listeners) is discussed below.

Figures 2 and 3 show the relative information transmission (Miller and Nicely, 1955) for the features of voicing, manner, and place of articulation for the auditory-only conditions (Fig.2) and for the visual-only and auditory-visual conditions (Fig. 3), respectively. In Fig. 2, information transmission for auditory place is lower than for either auditory

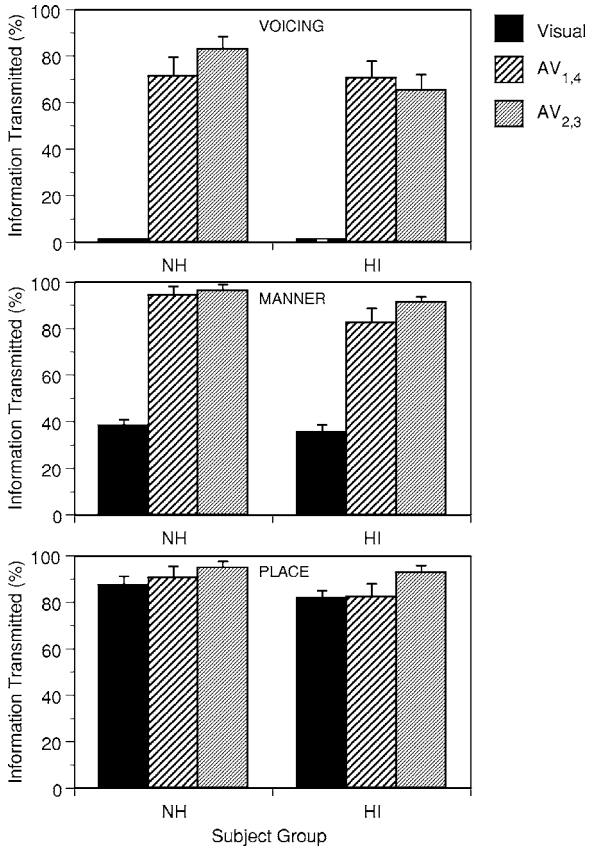

FIG. 3. Percent information transmitted in the three visual and auditory-visual conditions for the speech features voicing (top), manner-of-articulation (middle), and place-of-articulation (bottom). Data are for four normal-hearing (NH) subjects and four hearing-impaired (HI) subjects. Error bars are +1 standard error.

voicing or auditory manner in all conditions for both subject groups. This is consistent with place of articulation requiring greater acoustic bandwidth for optimum information transmission than either voicing or manner (Miller and Nicely, 1955). The two subject groups performed comparably in condition A_1 for all three features; however, the hearing-impaired subjects showed decrements in information transmission in most other conditions, most notably in the 2-, 3-, and 4-band conditions for the feature of place, in the two-band conditions for the feature of manner, and in the $A_{2,3}$ condition for the voicing feature. For the visual-only and auditory-visual conditions shown in Fig. 3, performance of the normal-hearing and hearing-impaired subjects was quite comparable. Voicing scores for both subject groups in the visual-only condition were nearly 0%, while mean manner scores were close to 40%, and place scores generally fell between 80% and 90%. This pattern is consistent with previous research described earlier (Grant and Walden, 1996; Grant *et al.*, 1998) showing that speechreading alone provides significant information about place of articulation, relatively little information about manner, and virtually no information about voicing. For the auditory-visual conditions, information transmission for voicing, manner, and place was similar across groups, despite decrements in auditory-alone performance of the hearing-impaired subjects (see, for example, the hearing-impaired subjects' gain in information transmission for the manner feature in conditions $AV_{1,4}$ and $AV_{2,3}$ compared with their performance in conditions $A_{1,4}$ and $A_{2,3}$). This finding is most likely related to the fact that good transmission of visual place information, combined with even limited voicing and manner information, will yield both high feature and high overall consonant recognition scores (Grant *et al.*, 1998; Christiansen *et al.*, in press).

Integration efficiency was calculated for the $A_{1,2,3,4}$,$AV_{1,4}$, and $AV_{2,3}$ conditions using Braida's PRE model, as described in the Methods section (Braida, 1991; Grant and Seitz, 1998). Recall that the integration of speech cues from different spectral regions or from different modalities is confounded by individual subject performance in the constituent conditions. In the case of the $A_{1,2,3,4}$ condition, the constituent conditions were $A_{1,4}$ and $A_{2,3}$. For the $AV_{1,4}$ condition, the constituent conditions were $A_{1,4}$ and V, and for the $AV_{2,3}$ condition, they were $A_{2,3}$ and V.

The PRE model analyzes the information received from the constituent conditions and combines the information in an optimal manner. The model then predicts the best possible recognition score that could be obtained given the subject's proficiency on the constituent conditions. For example, the model predicts the recognition scores obtained in condition $A_{1,2,3,4}$ by taking into account observed errors in conditions $A_{1,4}$ and $A_{2,3}$ and assuming perfect integration. Thus, it is possible to estimate the subject's skill at integrating information (as shown by the integration efficiency index), independent of the subject's ability to extract information (as shown by scores in the constituent conditions). This knowledge may be valuable for developing strategies for rehabilitation, whether they are based primarily on signal-processing schemes to aid the *extraction* of speech cues, or on training programs to improve the *integration* of speech cues.

The top panel of Fig. 4 shows the model predictions and observed scores for overall consonant recognition. Each point represents a single subject in one of the three predicted conditions, $A_{1,2,3,4}$, $AV_{1,4}$, and $AV_{2,3}$. Filled symbols show the data for normal-hearing subjects and unfilled symbols show results for hearing-impaired subjects. The diagonal line in the figure indicates equality between observed and predicted scores. Points above this line are cases where predicted scores are greater than obtained scores, and by definition, indicate sub-optimal integration. The largest deviations between predicted and observed scores occurred for hearing-impaired subjects, especially in the auditory-only condition $A_{1,2,3,4}$.

The bottom panel of Fig. 4 shows the average integration efficiency (observed/predicted * 100) for both groups of subjects. For normal-hearing subjects, integration efficiency was high ($>90\%$) regardless of whether the condition was auditory-only or auditory-visual. For hearing-impaired subjects, integration efficiency appears to be reduced, especially in the auditory-only condition $A_{1,2,3,4}$. A repeated-measures ANOVA was carried out with group as a between-subjects factor and condition as a within-subjects factor. The results showed a main effect for group $[F(1,6)=10.66, p=0.02]$, condition $[F(2,12)=11.27, p=0.002]$, and a group* condition interaction $[F(2,12)=13.33, p<0.001]$. However, when a similar analysis was conducted on just the auditory-visual

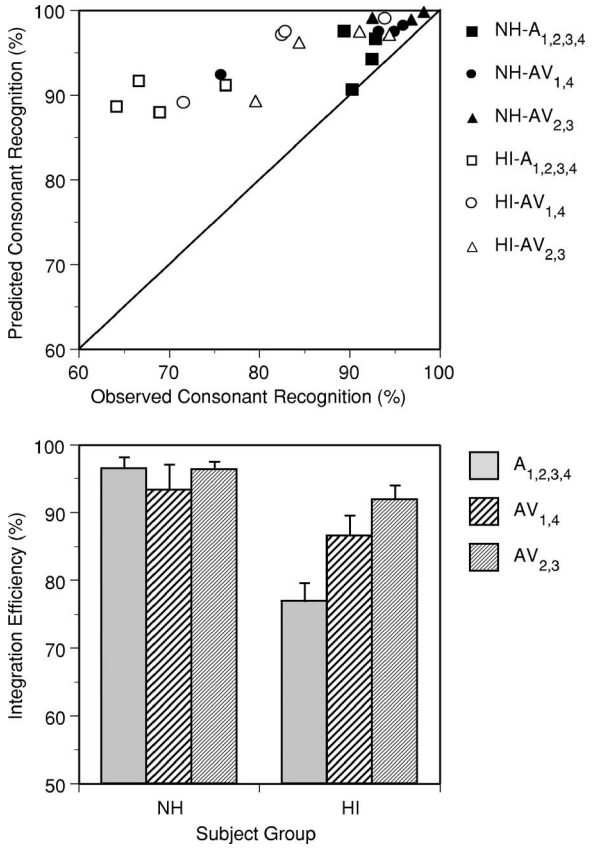

FIG. 4. (Top) Observed and predicted consonant recognition scores for $A_{1,2,3,4}$,$AV_{1,4}$, and $AV_{2,3}$ conditions. Scores for hearing-impaired (HI) subjects are shown by open symbols. The line indicates equality of observed and predicted scores. All scores to the left of this line indicate less than optimal integration. The greater the deviation from the diagonal line, the worse the subject is at integrating information from the different channels. (Bottom) Integration efficiency (IE) for NH and HI subjects. Integration efficiency is defined as the ratio between observed and predicted recognition performance. The difference in IE measures between NH and HI subjects was significant only for the $A_{1,2,3,4}$ condition. Error bars are +1 standard error.

conditions $AV_{1,4}$, and $AV_{2,3}$, none of the factors proved significant. This analysis indicates that the hearing-impaired subjects were able to integrate auditory-visual speech cues with the same relative efficiency as normal-hearing subjects, but had difficulty integrating auditory speech cues across spectral regions.

In interpreting the modeling results, the potential effects of band amplitude levels should be considered. Table III shows the RMS levels for each of the bands in each of the six auditory conditions. Recall that each of the first five auditory conditions was normalized in level to have the same overall rms amplitude as any other auditory condition. This normalization had the effect of altering the band levels depending on which bands were included in a given condition. For example, the levels of bands 1 and 4 were higher in the $A_{1,4}$ condition than they were in the $A_{1,2,3,4}$ condition. These differences in band levels are important for predictions made using the PRE model since the predictions are based on the information contained in the constituent conditions. Thus, if the band levels in the $A_{1,4}$ condition were higher than in the $A_{1,2,3,4}$ condition (as they were), then subjects could potentially extract more information from bands 1 and 4 when presented in the $A_{1,4}$ condition than from these same bands when presented in the $A_{1,2,3,4}$ condition, due to potential differences in audibility. This would lead to a model prediction that would exceed the actual scores by a greater amount and give the impression of reduced integration efficiency. To address this issue, the hearing-impaired subjects were re-tested on the original $A_{1,2,3,4}$ condition and on a new condition, $A_{1,2,3,4\text{sum}}$, which was created by adding the $A_{1,4}$ and $A_{2,3}$ signals together without subsequently normalizing the overall level. Because $A_{1,2,3,4\text{sum}}$ was not normalized, its overall amplitude was higher than $A_{1,2,3,4}$. Specifically, compared to $A_{1,2,3,4}$, this new signal had higher levels for band 1 (approximately 8 dB) and band 4 (approximately 8.5 dB), and roughly equal levels for bands 2 and 3 (see Table III). The subjects were tested first on ten blocks of 72 trials on the new $A_{1,2,3,4\text{sum}}$ condition, and then for three blocks each on the original $A_{1,2,3,4}$ condition and the $A_{1,2,3,4\text{sum}}$ condition, presented in random order. As expected, the results showed slightly better performance (roughly 5% absolute) for the more intense $A_{1,2,3,4\text{sum}}$ signal. However, this difference was not significant for any of the four hearing-impaired subjects. Therefore, it is unlikely that subjects' performance was affected by the differences in the specific magnitudes of the band levels.

IV. DISCUSSION

Normal-hearing and hearing-impaired subjects' performance on a consonant recognition task was evaluated under a variety of auditory, visual, and auditory-visual presentation conditions. Significant differences between the groups were observed in overall consonant recognition and feature transmission scores in the auditory-only conditions. However, performance was comparable for the two groups in the visual-only and auditory-visual conditions. Observed scores for a subset of the conditions were compared to predicted scores, using an optimum-processor model of integration. Differences between predicted and observed scores were used to estimate the efficiency with which subjects were able to integrate auditory cues across the frequency spectrum, or auditory and visual cues across sensory modalities.

With regard to place cues, hearing-impaired listeners extracted less information in the auditory-only conditions $A_{1,4}$ and $A_{2,3}$ than the normal-hearing listeners did (see the bottom panel of Fig. 2). However, place cues are transmitted very well visually, and hearing-impaired subjects performed as well as normal-hearing subjects in the visual-only condition, V (see the bottom panel of Fig. 3). Therefore, it is likely that the hearing-impaired listeners were able to overcome the deficit in auditory place information by making use of visual cues. Seen in this way, the similar performance of the groups for the place feature in conditions $AV_{1,4}$ and $AV_{2,3}$ in Fig. 3 is not surprising.

Unlike place cues, manner and voicing cues are not transmitted well visually. Voicing information, especially, is highly complementary with visual information and contributes greatly to auditory-visual speech recognition (Grant *et al.*, 1998). In the present study, the hearing-impaired listeners fared much worse at extracting voicing cues in condition $A_{2,3}$

than did the normal-hearing listeners (see the top panel of Fig. 2). Yet, in the $AV_{2,3}$ condition, which combines speechreading with $A_{2,3}$, voicing scores and overall consonant recognition were nearly comparable across groups. This finding may seem somewhat puzzling at first glance, in light of the expected adverse impact of limited voicing information on auditory-visual performance (Grant *et al.*, 1998; Erber, 2002). However, the features of voicing, manner and place are not independent of one another; good place identification, as in the visual-only condition, can aid in the identification of voicing and manner cues (Christiansen *et al.*, 2005; Christiansen *et al.*, in press).

Unlike their efficiency in auditory-visual integration, hearing-impaired listeners had difficulty combining information across widely separated audio-frequency channels, independent of their ability to extract information from these channels. This is shown in Fig. 4 by the reduced integration efficiency for condition $A_{1,2,3,4}$ compared to the normal-hearing subjects. One explanation for this finding concerns the possibility of upward spread of masking interfering with band 4. Some hearing-impaired listeners, especially those with broadened auditory filters and poor frequency selectivity, are known to exhibit excessive amounts of upward spread of masking (Gagné, 1988). It is possible that information contained in band 4 was partially masked when embedded in the $A_{1,2,3,4}$ condition. The PRE model cannot take into account such interactions that may occur between bands as a result of being presented concurrently. Instead, the model assumes that the information extracted from the $A_{1,4}$ condition has the full contribution of band 4 without any interference from mid-frequency auditory information. If masking of band 4 did occur, this could account for the fact that the hearing-impaired subjects obtained much lower scores than predicted in condition $A_{1,2,3,4}$ and, therefore, had poorer integration efficiency. The frequency bands used in this study were separated by an octave from one another in order to reduce such band-on-band interactions. However, given the broadened auditory filters that typically accompany sensorineural hearing loss, energy presented in band 3 may have provided enough masking to interfere with the phonetic information contained band 4.

Another possible explanation for the finding of reduced cross-spectral integration efficiency concerns the extent of the hearing loss in the vicinity of band 4 (4762–6000 Hz) and the possibility of "off-frequency" listening (i.e., responding to the energy in band 4 through the excitation of neurons tuned to lower frequencies where the hearing loss was not as great). Because the hearing-impaired subjects had extensive hearing loss in the band 4 region, some of the information in band 4 (e.g., low-intensity fricative energy) would have been inaudible (see Sec. II C for additional information). Based on the hearing thresholds shown in Table I, this would have been true for band 4 in the $A_{1,2,3,4}$ condition as well as in the $A_{1,4}$ condition. However, in condition $A_{1,4}$, there was no signal energy in frequency channels just below 4762 Hz (the low-frequency edge of band 4). It is possible that some of the information contained in band 4 could have been partially obtained by listening off-frequency, using less impaired auditory channels below 4762 Hz (Moore, 2004). When all four bands were presented in condition $A_{1,2,3,4}$, these same "off-frequency" channels may have been less able to extract the relevant information from band 4 because neurons tuned to the region just below band 4 were responding to frequency information associated with band 3, due to upward spread of masking. PRE model predictions of performance in condition $A_{1,2,3,4}$ were made under the assumption that all information extracted in the constituent conditions $A_{1,4}$ and $A_{2,3}$ was available for integration. This of course would not have been the case if the "off-frequency" listening hypothesis were correct. Thus, the poorer integration efficiency of the hearing-impaired listeners in condition $A_{1,2,3,4}$ may actually reflect an inability to extract and use high-frequency information when mid-frequency information is presented simultaneously.

This hypothetical scenario is not unlike previous suggestions (e.g., Doherty and Turner, 1996; Hogan and Turner, 1998) that hearing-impaired individuals have difficulty extracting and integrating high-frequency information from a broadband speech signal. More specifically, previous studies have claimed that high-frequency speech information may be useless for some hearing-impaired listeners with moderately severe to profound high-frequency hearing losses (e.g., Ching *et al.*, 1998; Hogan and Turner, 1998; Turner and Cummings, 1999). In particular, listeners with high-frequency "dead regions" in the cochlea (Moore, 2004) may be unable to make use of speech information that falls well within the dead region. The data in the present study conflict to a certain extent with these claims. The subjects in this experiment all had moderately severe to profound losses of between 60 and 105 dB HL in the region between 4000 and 6000 Hz (though they were not tested for the presence of dead regions). Yet, they were able to obtain selective benefit from high-frequency speech information, depending on the condition tested. In the studies previously referenced, high-frequency information was added to a broadband, lowpass-filtered speech signal. The lack of benefit seen when high-frequency information was added was interpreted to mean that listeners were unable to make any use of this information at all. Conditions $A_{1,2,3}$ and $A_{1,2,3,4}$ in the present study come closest to replicating the conditions in the previous studies. Indeed, limited benefit is seen by adding band 4 to bands 1, 2, and 3. However, the hearing-impaired listeners did benefit from band 4 when it was added to band 1 alone. This condition is unlike those in previous studies in that no mid-frequency information was presented along with band 4. Thus, amplified high-frequency speech information can be beneficial for listeners with significant high-frequency hearing loss, but only in contexts where no signal is presented in mid-frequency spectral regions.

More recently, studies have cast doubt on the notion that people with high-frequency hearing loss are unable to make use of amplified high-frequency speech information (Turner and Henry, 2002; Hornsby and Ricketts, 2003, 2006). In these studies, speech materials were presented in spectrally shaped noise or multitalker babble in various low-pass, high-pass or bandpass conditions. The high sound pressure levels required by the hearing-impaired listeners to ensure audibility may have reduced the utility of high-frequency speech

information to some extent. Nevertheless, the performance of the hearing-impaired listeners increased, albeit slightly in many cases, with the addition of high-frequency speech information. Listeners in these studies had hearing losses ranging up to 90 dB HL, but they were either not assessed for the presence of dead regions (Turner and Henry, 2002) or were found not to have dead regions (Hornsby and Ricketts, 2003, 2006). Certainly, it appears reasonable that if high-frequency information is presented to a dead region, then this information is only beneficial insofar as other, less damaged regions can make use of it. However, the results of these studies and the present study, which show that hearing-impaired listeners can benefit from high-frequency speech information in selected conditions, suggest caution with respect to limiting high-frequency amplification for listeners based solely on audiometric thresholds or audiometric configuration.

A third, and perhaps most likely, explanation for the hearing-impaired listeners' reduced cross-spectral integration efficiency may be related to an explanation suggested by Turner and Henry (2002) as to why their hearing-impaired subjects were able to make use of high-frequency speech information while other studies showed no benefit. In their study, consonant recognition performance was relatively poor, even in the broadest bandwidth condition, due to the presence of multitalker babble. The addition of high-frequency information to the various low-pass filter conditions may have provided listeners with enough additional information to decipher the "easier" features of speech, such as voicing. In earlier studies of speech in quiet (e.g., Hogan and Turner, 1998), much of the speech signal was already audible to the hearing-impaired listeners. Thus, additional high-frequency information was not as helpful to these listeners in deciphering more "difficult" features, such as place of articulation. In the present study, overall consonant recognition scores for band 1 alone were low ($<30\%$). Analogous to the results of Turner and Henry (2002), the addition of band 4 to band 1 may have boosted scores considerably by providing information related to easier speech cues, whereas the addition of band 4 to bands 1, 2, and 3 may not have provided enough additional information to allow listeners to resolve the remaining, more difficult speech cues. As with the previous two explanations, an inability to make use of speech information in band 4 when it was presented concurrently with bands 1, 2, and 3 would give the appearance of reduced integration efficiency.

The reduced cross-spectral integration efficiency of the hearing-impaired listeners may not simply be a product of confounding the extraction and integration of information, but may in fact be evidence of a real deficit in the across-frequency processing of temporal information. As described in the Introduction, Healy and Bacon (2002) found that listeners with sensorineural hearing loss were less successful than normal-hearing listeners in integrating temporal speech information across two widely separated frequency channels. This finding is especially convincing, since their stimuli were designed to eliminate other confounding effects of sensorineural hearing loss such as broadened auditory filters and audibility.

Although age effects were not specifically assessed in this study, it is worthy of mention that the hearing-impaired subjects in the present study were all over 65 years of age, while the normal-hearing subjects were all 55 years of age and younger. Previous investigators (e.g., Spehar *et al.*, 2004; Sommers *et al.*, 2005b) have reported poorer consonant and word recognition in visual-only conditions for older versus younger adults. In contrast, the visual-only performance of older (hearing-impaired) and younger (normal-hearing) adults in the present study was comparable. This contrasting finding may be a reflection of differences in the ages of the "younger" subjects in each study. In the studies previously referenced, the younger adults were in their early- to mid-twenties, while those in the present study ranged in age from 29 to 55 years.

The auditory-visual integration abilities of older versus younger adults have been assessed using subject groups that were matched either for peripheral hearing sensitivity (Cienkowski and Carney, 2002, 2004) or for auditory-only performance (Sommers *et al.*, 2005b). Despite their poorer speechreading abilities, older adults' auditory-visual integration abilities appeared to be comparable to those of younger adults. Auditory-visual integration was assessed by analyzing responses to variations of a McGurk paradigm (Cienkowski and Carney, 2002, 2004) or by calculating measures of auditory and visual enhancement similar to the Sumby and Pollack (1954) relative benefit measure (Sommers *et al.*, 2005b). As discussed previously, without taking into account individual phoneme confusions in the unimodal conditions, such measures may not provide an accurate assessment of listeners' integration abilities. Nevertheless, the findings of these studies do corroborate the indirect finding in the present study of no age effects in auditory-visual integration ability.

Finally, experiments with time-compressed *auditory* speech have revealed age effects suggestive of a slowing in central auditory processing (e.g., Gordon-Slant and Fitzgibbons, 1999). However, Spehar *et al.* (2004) found no such age effects for temporally altered *visual* speech signals. Spehar *et al.* (2004) finding suggests that the slowing of central processing with age may be modality-specific. While not conclusive in and of itself, the finding does leave open the possibility that the normal cross-modal but abnormal cross-spectral integration observed in the hearing-impaired subjects may have been partially due to differences in the effects of age on the speed of central processing across modalities. However, a recent study by Souza and Boike (2006) suggests that age may not be a significant factor in cross-spectral integration, at least with respect to speech signals composed of mostly temporal cues. Specifically, these authors examined the effect of age on the ability to combine temporal-envelope information across frequency channels. They assessed consonant identification in 1-, 2-, 4-, and 8-channel conditions processed to restrict spectral cues, as well as in an unprocessed condition. Results revealed a significant trend toward poorer performance with increasing age in all conditions except the unprocessed condition, suggestive of an age-associated deficit in the use of temporal-envelope information. However, no age-associated deficit was observed in combining information across frequency channels. There-

fore, it seems unlikely that age differences between the normal-hearing and hearing-impaired subjects in this study contributed significantly to the differences seen in their cross-spectral integration abilities.

Aural rehabilitation options for improving speech recognition should exploit the good use of visual cues that the hearing-impaired individuals were able to make. Despite poorer auditory-only performance compared with normal-hearing subjects, listeners with hearing loss were able to use the visual speech signal to overcome these deficits and achieve normal auditory-visual consonant recognition. Therefore, aural rehabilitation efforts that emphasize environmental and behavioral alterations to make better use of visual speech cues are likely to greatly aid the hearing-impaired listener in understanding speech, particularly in noisy or reverberant conditions. Furthermore, this study and previous studies have shown that hearing-impaired listeners are able to benefit from the provision of high-frequency speech information in certain impoverished audio conditions (e.g., due to noise or to spectral filtering). These results suggest caution with respect to limiting high-frequency amplification for listeners based solely on audiometric thresholds or audiometric configuration.

V. CONCLUSIONS

Comparisons of within-modality (auditory-only) and across-modality (auditory-visual) integration efficiency demonstrated that both normal-hearing and hearing-impaired subjects had little trouble integrating auditory-visual information. However, hearing-impaired listeners showed difficulty integrating spectral information across widely separated audio frequency channels. Specifically, high-frequency speech information (4762–6000 Hz) was found to be useful when combined with low-frequency speech information (298–375 Hz), but was less useful when mid-frequency speech information (1890–2381 Hz) was also present. These results support the contention that amplified high-frequency speech information can be beneficial for listeners with significant high-frequency hearing loss in quiet, particularly in contexts where no signal is presented in adjacent lower-frequency spectral regions. Future tests comparing the information content of the different spectral bands should help further our understanding of the auditory integration difficulties observed in hearing-impaired subjects. In the absence of clear evidence contraindicating the amplification of high-frequency speech information, such amplification combined with the use of visual cues is recommended to improve speech understanding for individuals with severe high-frequency sensorineural hearing loss in adverse listening conditions.

VI. ACKNOWLEDGMENTS

This research was supported by the Clinical Investigation Service, Walter Reed Army Medical Center, under Work Unit #00-2501, Grant No. DC 00792 from the National Institute on Deafness and Other Communication Disorders, Grant No. SBR 9720398 from the Learning and Intelligent Systems Initiative of the National Science Foundation to the International Computer Science Institute, and the Oticon Foundation, Copenhagen, Denmark. We would also like to thank Rosario Silipo for assistance in creating the auditory stimuli, Mary Cord for assistance in data collection, and Louis Braida for his help fitting the Prelabeling Model of Integration to our data. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

American National Standards Institute (**1989**). Specifications for audiometers (ANSI S3.6-1989). (ANSI, New York).

Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (**1989**). "Speech perception using combinations of auditory, visual, and tactile information," J. Rehabil. Res. Dev. **26**, 15–24.

Borg, I., and Lingoes, J. (**1987**). *Multidimensional Similarity Structure Analysis* (Springer-Verlag, New York).

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. **43**, 647–677.

Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Christiansen, T. U., Dau, T., and Greenberg, S. (in press). "Spectro-temporal processing of speech–An information-theoretic framework," 14th International Symposium on Hearing.

Christiansen, T. U., and Greenberg, S. (**2005**). "Frequency selective filtering of the modulation spectrum and its impact on consonant identification," 21st Danavox Symposium: Hearing Aid Fitting, **21**, 585–599.

Cienkowski, K. M., and Carney, A. E. (**2002**). "Auditory-visual speech perception and aging," Ear Hear. **23**, 439–449.

Cienkowski, K. M., and Carney, A. E. (**2004**). "The integration of auditory-visual information for speech in older adults," J. Speech Lang. Path. Aud. **28**, 166–172.

Doherty, K. A., and Turner, C. W. (**1996**). "Use of a correlational method to estimate a listener's weighting function for speech," J. Acoust. Soc. Am. **100**, 3769–3773.

Erber, N. P. (**2002**). "Hearing, vision, communication, and older people," Seminars in Hearing **23**, 35–42.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Gagné, J. P. (**1988**). "Excess masking among listeners with a sensorineural hearing loss," J. Acoust. Soc. Am. **83**, 2311–2321.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.

Grant, K. W. (**2002**). "Measures of auditory-visual integration for speech understanding: A theoretical perspective (L)," J. Acoust. Soc. Am. **112**, 30–33.

Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Greenberg, S., and Arai, T. (**2004**). "What are the essential cues for understanding spoken language?" IEICE Trans. Inf. Syst. **87D**, 1059–1070.

Greenberg, S., Arai, T., and Silipo, R. (**1998**). "Speech intelligibility from exceedingly sparse spectral information," *Proc. Int. Conf. Spoken Lang. Processing*, Sydney, Australia, 74–77.

Healy, E. W., and Bacon, S. P. (**2002**). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1262–1275.

Healy, E. W., Kannabiran, A., and Bacon, S. P. (**2002**). "An across-frequency processing deficit in listeners with hearing impairment is sup-

ported by acoustic correlation," J. Speech Lang. Hear. Res. **48**, 1236–1242.
Hill, F. J., McRae, L. P., and McClellan, R. P. (**1968**). "Speech recognition as a function of channel capacity in a discrete set of channels," J. Acoust. Soc. Am. **44**, 13–18.
Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.
Hornsby, B. W. Y., and Ricketts, T. (**2003**). "The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding," J. Acoust. Soc. Am. **113**, 1706–1717.
Hornsby, B. W. Y., and Ricketts, T. (**2006**). "The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. II. Sloping hearing loss," J. Acoust. Soc. Am. **119**, 1752–1762.
Kruskal, J. B., and Wish, M. (**1978**). *Multidimensional Scaling* (Sage, Beverly Hills, CA).
Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). "Vowel and consonant resolution: Basic sensitivity and context memory," J. Acoust. Soc. Am. **84**, 1262–1280.
Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Erlbaum Hillsdale, NJ).
Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).
Massaro, D. W., and Cohen, M. M. (**2000**). "Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception," J. Acoust. Soc. Am. **108**, 784–789.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**2004**). "Dead regions in the cochlea: Conceptual foundations, diagnosis, and clinical applications," Ear Hear. **25**, 98–116.
Müsch, H., and Buus, S. (**2001**). "Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920.
Pickett, J. M. (**1999**). *The Acoustics of Speech Communication* (Allyn and Bacon, Boston, MA).
Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (**2006**). "Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments," Cerebral Cortex Advance Access, published June 19, 2006 at http://cercor.oxfordjournals.org/cgi/reprint/bhl024v1
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Silipo, R., Greenberg, S., and Arai, T. (**1999**). "Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations," *Proc. Eurospeech*, Budapest, Hungary, 2687–2690.
Sommers, M. S., Spehar, B., and Tye-Murray, N. (**2005a**). "The effects of signal-to-noise ratio on auditory-visual integration: Integration and encoding are not independent (A)," J. Acoust. Soc. Am. **117**, 2574.
Sommers, M. S., Tye-Murray, N., and Spehar, B. (**2005b**). "Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults," Ear Hear. **26**, 263–275.
Souza, P. E., and Boike, K. T. (**2006**). "Combining temporal-envelope cues across channels: Effects of age and hearing loss," J. Speech Lang. Hear. Res. **49**, 138–149.
Spehar, B., Tye-Murray, N., and Sommers, M. (**2004**). "Time-compressed visual speech and age: A first report," Ear Hear. **25**, 565–572.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Turner, C. W., and Cummings, K. J. (**1999**). "Speech audibility for listeners with high-frequency hearing loss," Am. J. Audiol. **8**, 47–56.
Turner, C. W., and Henry, B. A. (**2002**). "Benefits of amplification for speech recognition in background noise," J. Acoust. Soc. Am. **112**, 1675–1680.
Turner, C. W., Chi, S. L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.
Wang, M. (**1976**). "SINFA: Multivariate uncertainty analysis for confusion matrices," Behav. Res. Methods Instrum. **8**, 471–472.