Effects of consonant-vowel intensity ratio on loudness of monosyllabic words

Suzanne B. Orr and Allen A. Montgomery[a)]
Department of Communication Sciences and Disorders, The Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208

Eric W. Healy
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

Judy R. Dubno
Department of Otolaryngology—Head and Neck Surgery, Medical University of South Carolina, Charleston, South Carolina 29425

(Received 2 December 2009; revised 20 August 2010; accepted 25 August 2010)

Previous research has suggested that speech loudness is determined primarily by the vowel in consonant-vowel-consonant (CVC) monosyllabic words, and that consonant intensity has a negligible effect. The current study further examines the unique aspects of speech loudness by manipulating consonant-vowel intensity ratios (CVRs), while holding the vowel constant at a comfortable listening level (70 dB), to determine the extent to which vowels and consonants contribute differentially to the loudness of monosyllabic words with voiced and voiceless consonants. The loudness of words edited to have CVRs ranging from −6 to +6 dB was compared to that of standard words with unaltered CVR by 10 normal-hearing listeners in an adaptive procedure. Loudness and overall level as a function of CVR were compared for four CVC word types: both voiceless consonants modified; only initial voiceless consonants modified; both voiced consonants modified; and only initial voiced consonants modified. Results indicate that the loudness of CVC monosyllabic words is not based strictly on the level of the vowel; rather, the overall level of the word and the level of the vowel contribute approximately equally. In addition to furthering the basic understanding of speech perception, the current results may be of value for the coding of loudness by hearing aids and cochlear implants.
 [DOI: 10.1121/1.3493426]

PACS number(s): 43.71.Es, 43.71.An, 43.66.Cb [MS] Pages: 3105–3113

I. INTRODUCTION

There is reason to believe that the perception of loudness in speech has unique elements beyond the obvious influence of overall sound pressure level (SPL). While the relationship between signal intensity and the perceptual correlate of loudness has been relatively well defined for pure tones (Stevens, 1955, 1956), far less is known about this relationship for speech signals. This uniqueness of speech loudness first became apparent in a study by Lehiste and Peterson (1959), in which isolated vowels and consonant-vowel-consonant (CVC) words were recorded as produced by one male talker under conditions of constant talker effort, allowing natural variation in intensity. The result involving isolated vowels was a set of 10 utterances *that sounded equally loud*, even though the acoustic energy in the tokens differed by as much as 5.3 dB.

Note that 5 dB in the conversational range of intensity (40–80 dB SPL) is a large, detectible difference, well in excess of the 1.3 dB difference limen for intensity of words (Rogers *et al.*, 2006). That is, if intensity were the only basis for speech loudness, the words would not have sounded equally loud. Because the talker's effort was constant, it was concluded that the effort a talker puts into producing a word is important in the perception of the loudness of the word.

Subsequently, a large amount of research in speech loudness has involved the concept of "talker effort" or "vocal effort." Much of this work has considered that talker effort, as experienced and associated with acoustic energy level, and internalized by the talker, is an important factor in judging speech loudness (Ladefoged and McKinney, 1963; Allen, 1971), although other, purely acoustic correlates of increasing vocal level influence speech loudness as well (Brandt *et al.*, 1969; Mendel *et al.*, 1969). One assumption has been that subglottic pressure (SGP) during the production of the vowel, measured in cm H_2O and sensed mainly in the larynx and trachea, is a primary measure of vocal effort.

It is also possible, however, that sonorant consonants with vocalic formant structures (such as liquids, nasals, and glides), and their associated SGPs, contribute to the perception of vocal effort and thus may influence judgments of speech loudness. Furthermore, the muscular tension associated with voicing itself may contribute to the perceptual basis of vocal effort, thus producing a possible differential effect of voicing on speech loudness (Allen, 1971). These studies raise the question of whether the vowel and consonant have differential influences on the loudness of speech.

[a)]Author to whom correspondence should be addressed. Electronic mail: amontgom@mailbox.sc.edu

In particular, it is possible that sonorant consonants, voiceless (continuant) consonants, and vowels all make different contributions to the loudness of CVC monosyllabic words.

To our knowledge, only one study has explored the extent to which the intensity of the consonants in CVC monosyllabic words influences the perception of their loudness, or whether information about speech loudness is contained primarily in the vowel (Montgomery *et al.*, 1987). The findings supported "vowel dominance" and suggested that consonant intensity has a negligible effect on loudness judgments. In that study, the intensity of one consonant (initial or final) in four CVC monosyllabic words ("seen," "gave," "thin," and "robe") was scaled in order to modify the consonant-vowel intensity ratio (CVR), resulting in a set of six tokens (−6, −3, 0, +3, +6, and +9 dB CVR) plus the unaltered stimulus. Normal-hearing and hearing-impaired listeners participated in a loudness matching task, with the standard word ("laugh") presented at 90 dB SPL, and were instructed to adjust the attenuator dial until the standard word and token sounded equally loud. The functions relating loudness to CVR were plotted and the slopes of the functions were calculated. The slopes of the function obtained for each word by normal and hearing-impaired listeners were near zero, with small standard deviations, meaning that no consistent change in loudness was observed when the vowel was maintained at a constant level. Whether this result obtained with normal-hearing and hearing-impaired listeners at a high level for the fixed standard word and limited consonantal context holds for a wider range of monosyllabic words presented at a more comfortable overall level remains to be explored.

The present study sought to further our understanding of the sources of speech loudness by manipulating CVRs (holding the vowel intensity constant) to determine the extent to which vowels and consonants contribute differentially to the loudness of monosyllabic CVC words with voiced (V) and voiceless (VL) consonants. The primary outcomes of this study were the slopes of the functions relating loudness to CVR for four word types: (1) CVC words with both voiceless consonants modified (VL); (2) CVC words with only the initial voiceless consonants modified and the final voiceless consonants remaining at their unaltered level (VL-I); (3) CVC words with both voiced consonants modified (V); and (4) CVC words with only the initial voiced consonants modified and the final voiced consonants remaining at their unaltered level (V-I). The experimental questions involve a comparison of the slopes and the linear or nonlinear nature of the functions derived from the four word types across five CVRs. In contrast to the Montgomery *et al.* (1987) study in which only one of the two consonants was altered, in the present study both initial and initial plus final consonants were altered, in order to examine further the effects of overall level on loudness. Further, the use of stimuli with both voiced and voiceless consonants allows the influence of voicing with its attendant SGP and laryngeal tension to be examined in contrast with voiceless consonants.

TABLE I. List of CVC stimulus words with voiceless (VL) or voiced (V) consonantal contexts. VL-I and V-I indicate the words where only the initial consonant was altered in consonant-to-vowel ratio.

Voiceless (VL)	Voiceless-initial (VL-I)	Voiced (V)	Voiced-initial (V-I)
Chaff	Chef	Kneel	Lawn
Chef	Church	Lamb	Man
Church	Fish	Lawn	Numb
Fish	Sis	Learn	Room
Fuss	Thief	Limb	Rum
Sash		Man	
Sauce		Moon	
Sis		Numb	
Surf		Room	
Thief		Rum	

II. METHOD

A. Participants

Participants were 10 individuals ranging in age from 22–69 years (8 whose age was <38 years and 2 participants aged 66 and 69 years) having audiometric thresholds below 20 dB HL for octave frequencies between 250 and 4000 Hz. All were native speakers of American English and had previous experience in psychoacoustic tasks. Listeners not connected with the laboratory were paid for their participation.

B. Stimulus preparation

Productions of 20 CVC monosyllabic words were recorded in an audiometric booth by an articulate male speaker with a standard American dialect using a Shure SM57 microphone. The productions were amplified using a Mackie 1202-VLZ mixer and were digitized on a PC at 44.1 kHz with 16-bit resolution using an Echo Gina 24 D/A converter. Ten words contained only voiceless fricatives or affricates in order to allow the vowel to be the only voiced segment of the word. Ten words contained only voiced consonants (nasals and liquids), which have formant structures and SGPs similar to vowels (see Table I).[1] First, the unaltered CVR was calculated for each word as the difference in dB between the RMS average of the entire initial or final consonant and the middle 100 ms of the vowel.[2] Judgments of C-V boundaries, especially in the case of voiced consonants, were made with the aid of spectrograms and repeated listening to longer and shorter sub-segments of the waveform. The unaltered stimuli had CVRs that averaged −7.4 dB, and ranged from −21.2 dB for the initial consonant in "thief" to −2.4 dB for that in "sash." The mean unaltered CVRs for each class of phonemes were as follows: fricative (initial/final consonant)= −9.3/−11.4, affricate=−4.7/−9.5, sonorant=−5.6/−4.7 dB.

These CVRs determined the amount of amplification or attenuation required for each consonant to achieve the experimental CVRs. The experimental CVRs (equal in level for initial and final consonants) ranged from −6 to +6 dB in 3-dB steps, and were set by amplifying or attenuating the consonant portions of the waveform using a waveform editor. The vowel portions of the waveforms were not modified. In addition to these alterations in which initial and final consonants were modified to produce the same CVRs, the CVRs

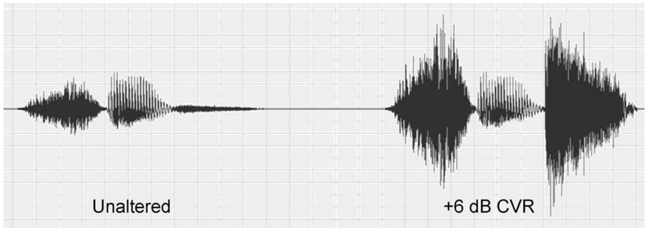

FIG. 1. Waveforms of the CVC word "chef" in unaltered state and modified to produce a +6 dB consonant-vowel intensity ratio (CVR).

of only the initial consonants of 5 words with voiced consonants and 5 words with voiceless consonants were also edited from −6 to +6 dB while leaving the CVR of the final consonants unaltered. This resulted in 20 CVC words, 10 of which had voiced consonants and 10 of which had voiceless consonants in which CVRs of both initial and final consonants were modified (V and VL), and a subset of 5 words (VL-I) and 5 words (V-I) in which the CVRs of only the initial consonants were modified. Modifications were performed at zero crossings, and considerable care was taken to ensure that the stimuli were altered only in the manner intended. The resulting stimuli were screened carefully for editing artifacts by three experienced listeners who employed repeated presentations at comfortably loud listening levels over Telephonics TDH-39P headphones. Figure 1 shows an example of a word in its unaltered and modified states.

C. Stimulus properties

Table II shows that the mean duration of the voiced initial consonants was 85.5 ms, compared to the voiceless initial consonant duration of 162.4 ms. Note that the total mean duration of the vowel and final consonants is similar for words with VL and V consonants (349.1 and 381.8 ms, respectively). Further, the words with voiced consonants showed the typical influence of voicing on vowel duration, increasing from 153.6 ms (VL) to 213.4 ms (V) (Luce and Charles-Luce, 1985; Summers, 1987).

Figure 2 displays the overall intensity of the stimuli with various CVRs. As expected, the stimuli containing voiced

TABLE II. Means and standard deviations of durations (in ms) of initial consonants (IC), vowels (VO), and final consonants (FC) for words with voiceless (VL) and voiced (V) consonants.

Word type VL	IC	VO	FC	Word type V	IC	VO	FC
Chaff	134	201	168	Kneel	94	186	208
Chef	193	135	182	Lamb	94	253	145
Church	110	143	202	Lawn	65	286	150
Fish	143	109	229	Learn	95	238	146
Fuss	151	139	213	Limb	91	137	182
Sash	180	212	184	Man	63	271	175
Sauce	181	218	191	Moon	92	235	157
Sis	208	117	231	Numb	99	193	165
Surf	180	117	180	Room	76	131	183
Thief	144	145	175	Rum	86	204	173
Mean	162.4	153.6	195.5		85.5	213.4	168.4
SD	30.5	41.1	22.3		12.9	52.8	19.9

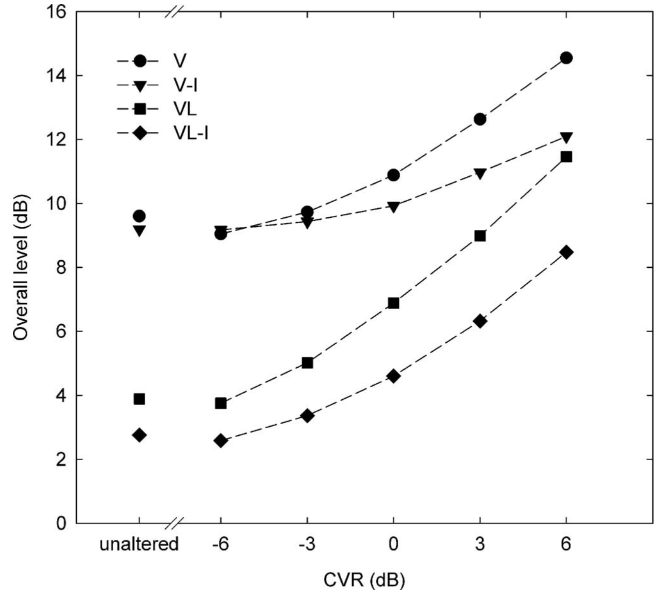

FIG. 2. Overall level (in arbitrary-reference dB values) as a function of altered consonant-vowel intensity ratio (CVR) in CVC stimuli with constant vowel intensity.

consonants are more intense at all CVRs than those with voiceless consonants (Blood, 1981). Because the level of the vowel was held constant, the overall intensities of the stimuli vary only as a function of the level of the initial, or initial and final consonants. It may be seen that the stimuli containing voiceless consonants show a separation of approximately 2 dB between stimuli in which both initial and final consonants were amplified and those in which only the initial consonant was amplified. The words containing voiced consonants show a similar separation except at −6 and −3 dB CVR, where initial and initial-plus-final consonant modifications produced similar overall intensities. This is because the brief initial voiced consonants contribute little to overall level at these low CVRs. Also note that the increase in overall intensity with increasing CVR (from −6 to +6 dB) for the stimuli with voiced consonants is much less than the corresponding change for the stimuli with voiceless consonants.

Table III includes linear and quadratic trends for the slopes of functions relating overall level to CVR in Fig. 2. Note that, in addition to the basic linear relationship between overall level and CVR, the slopes for VL, VL-I, and V had a significant quadratic component and the slope for V-I approaches significance for the quadratic trend. The obvious quadratic component in these curves is due to the logarithmic nature of the modifications to the consonants. That is, the −6 dB CVR consonants contribute little to the overall level, whereas the same consonants at +6 dB CVR contribute significantly. Thus, the functions relating the overall levels of the words to CVR show a strong linear component and a quadratic component. Whether this nonlinear property of the stimuli influences the loudness judgments of the listeners becomes an empirical question.

D. Procedure

Participants were seated comfortably in an audiometric booth and were administered CVC monosyllabic words in

TABLE III. Trend analysis of linear and quadratic components of the functions relating overall level to consonant-vowel intensity ratio (CVR) for four word types. Probabilities were considered significant at $\alpha=0.01$ and are displayed in bold.

Word type	Linear/quadratic	F	Probability
VL	Lin	679.195	**<0.001**
	Quad	172.195	**<0.001**
VL-I	Lin	274.415	**<0.001**
	Quad	3420.623	**<0.001**
V	Lin	234.810	**<0.001**
	Quad	36.873	**<0.001**
V-I	Lin	39.582	**0.003**
	Quad	18.219	0.013

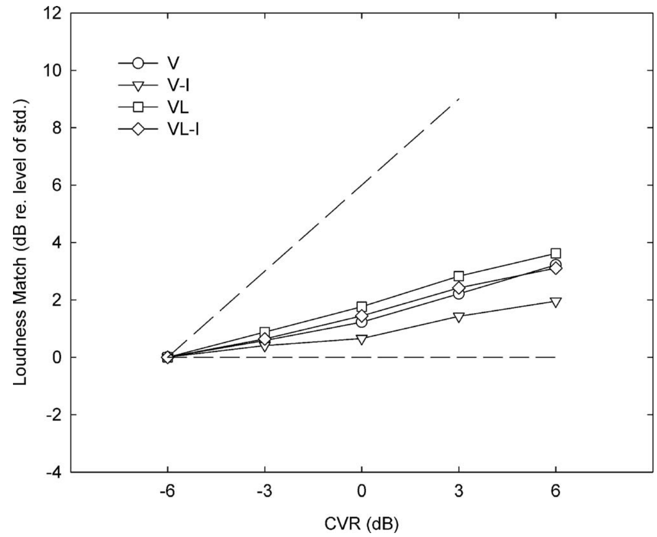

FIG. 3. Loudness match (in dB relative to the level of the standard) as a function of consonant-vowel intensity ratio (CVR) in CVC stimuli with constant vowel intensity. Standard deviations for each mean plotted here ranged from 0.8 to 1.4. Dashed reference lines for slopes of 0.0 and 1.0 are provided.

which the altered CVR order was randomized. The digital files were played from a PC (Edirol UA-5 D/A converter) and presented binaurally through Telephonics TDH-39P headphones. The output levels of the headphones were calibrated using a Larson Davis model 800B sound level meter with a Larson Davis AEC101 6 cc coupler and a 1-in. Bruel & Kjaer microphone. To calibrate the presentation levels of the unaltered and associated altered stimuli, the unaltered production of each word was set to 70 ± 2 dBA at each earphone (peak of the fast-response RMS average). In each loudness match trial (see below), the standard was always the unaltered form of the word and the stimulus was always that same word with altered CVR, but with a vowel intensity identical to that of the standard.

Loudness matches were obtained using an adaptive two-interval, forced choice paradigm, based on a procedure used by Florentine *et al.* (1996), with a one-up, one-down decision rule to track the 50% point on the psychometric function relating loudness to level (Levitt, 1971). The unaltered CVC monosyllabic word (standard) was presented, followed by 500 ms of silence, and then the same word with altered CVR (stimulus word). The listener was asked to indicate whether the stimulus was louder or softer than the standard by pressing a key, completing one trial. If the stimulus was judged louder, the overall intensity of the stimulus was reduced in the next paired presentation. If the stimulus was judged softer, the level of the stimulus was increased in the subsequent presentation. The presentation started with the intensity of the stimulus noticeably greater than that of the standard, with the starting difference randomly adjusted for each word over a range of 6 dB. A step size of 3 dB was used until after the second reversal, whereupon it was reduced to 1 dB. The procedure was terminated after nine reversals (Silva and Florentine, 2006). The average level in dB of the last six reversals was used to estimate the level of the stimulus judged to be equally loud to the standard, constituting a run for that CVR for that word. Mean loudness matches in dB were calculated for each of three such adaptive runs at each CVR for each stimulus word. In the subsequent figure, the loudness match is reported in dB relative to the standard level.

The nature of this experiment required that participants be highly consistent in their loudness matching ability. To this end, each listener completed a training period of at least two hours, during which they completed multiple runs involving practice words with both VL and V consonants at 5 CVRs, from −6 to +6 dB. At the end of training, all listeners consistently produced loudness matches within 2 dB across three runs of a given CVR and standard deviations of reversal levels less than 2 dB within a run. During the test procedure, participants were instructed to repeat a run once if the loudness match within a CVR differed by more than 2 dB across runs or if the standard deviation of reversals within a run exceeded 2 dB. In this case, the mean of all four runs was calculated for the analysis. Listeners participated in two-hour increments and required 12–14 h to complete the experiment.

III. RESULTS

The primary data obtained in this study are related to the effect of the altered speech material on the perception of speech loudness. The consonants of CVC stimuli were modified in intensity to observe the effects on speech loudness, yielding functions relating loudness match to CVR, with CVRs ranging from −6 to +6 dB. The main focus of the data analysis was to compare the slopes of these “loudness functions” among word types and in relation to comparable slopes reflecting overall stimulus levels (shown in Fig. 2).

The mean loudness matches for the four word types at the 5 CVRs were calculated for 10 listeners and are seen in Fig. 3. Loudness estimates are based on a total of approximately 4600 runs. A dashed line with a zero slope illustrates the possible result of no change in loudness with increasing CVR, indicating that only vowel level contributed to loudness (vowel dominance), given that vowel level was held constant. A dashed line with a slope of 1 is also included for reference. The slopes of the loudness functions for each word

TABLE IV. Slopes of functions relating loudness to consonant-vowel intensity ratio (CVR) and functions relating overall level to CVR for four word types.

	Loudness					Overall level		
			99% Confidence interval					
Word type	Mean slope	SD	Lower bound	Upper bound	r^2	Mean slope	SD	r^2
VL	0.306	0.062	0.244	0.368	0.984	0.646	0.078	0.987
VL-I	0.267	0.058	0.208	0.326	0.995	0.491	0.066	0.967
V	0.269	0.055	0.211	0.325	0.983	0.464	0.096	0.966
V-I	0.164	0.039	0.123	0.203	0.970	0.246	0.088	0.934

type, averaged across words and listeners, are shown in Table IV. The mean slopes for the overall level for the four word types (seen in Fig. 2) were also calculated and are also shown in Table IV.[3]

Despite the general linear appearance of the loudness functions in Fig. 3, it is necessary from both a practical and theoretical point of view to determine whether there is a quadratic component present in the functions. This is theoretically important because a marked increase in loudness when the CVR exceeds zero would indicate that the vowel is dominant at lower levels, with a larger contribution of the consonant at higher levels. An alternate explanation is that the listeners used nonlinear growth in overall level as one of the bases of loudness. Additionally, such nonlinearity would limit the use of linear loudness slopes as the basis for analysis. Table V shows the quadratic trend for the loudness functions of the four word types. Note that the quadratic component is significant only for stimuli with voiced consonants. However, inspection of Fig. 3 indicates minimal quadratic influence on the shape of the curves. Furthermore, the ratio of linear to quadratic mean squares was always greater than 50 to 1, indicating the predominance of the linear component. Hence, the loudness data employed in all subsequent analyses involve slopes based on linear curve fitting under various experimental conditions.

TABLE V. Trend analysis of linear and quadratic components of the functions relating loudness to consonant-vowel intensity ratio (CVR) for the four word types. Probabilities were considered significant at $\alpha=0.01$ and are displayed in bold.

Word type	Linear/quadratic	F	Probability
VL	Lin	213.302	**<0.001**
	Quad	0.889	0.370
VL-I	Lin	241.307	**<0.001**
	Quad	0.005	0.946
V	Lin	240.789	**<0.001**
	Quad	30.765	**<0.001**
V-I	Lin	178.230	**<0.001**
	Quad	5.154	0.049

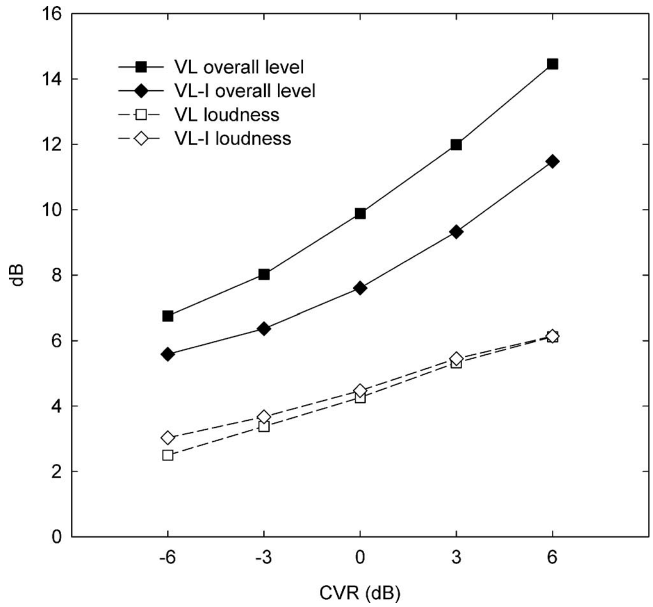

FIG. 4. Comparison of loudness and overall level (in arbitrary-reference dB values) as a function of consonant-vowel intensity ratio (CVR) for stimuli with both voiceless (VL) consonants and stimuli with only the initial voiceless consonant (VL-I) altered. Functions relating loudness and overall level to CVR are separated arbitrarily for ease of viewing.

The slopes of the functions relating loudness to CVR for the four word types were compared using a repeated measures ANOVA adjusted for violations of sphericity with Greenhouse-Geisser adjustments, $F(3,27)=17.61$, $p<0.001$. Post hoc comparisons with Bonferroni adjustment $\alpha=0.008$ (overall $\alpha=0.05,6$ comparisons) indicated that the slopes of VL (0.306), VL-I (0.267), and V (0.269) were not significantly different from one another, but the slope of V-I (0.164) was significantly lower than the others. Furthermore, as seen in the 99% confidence intervals, the loudness slopes for all four word types based on these estimates are greater than 0.0 and less than 1.0 and are statistically significantly lower than the corresponding slopes for the functions relating overall levels of the stimuli to CVR (see Table IV).

Following analysis of the loudness functions, a subsequent analysis of the slopes of the functions relating overall level to CVR (Fig. 2), using the significantly linear slopes for comparison to the linear slopes of the functions relating loudness to CVR, yielded virtually identical results. Following a significant main effect of word type, $F(3,26)=25.5$, $p<0.001$, post hoc comparisons with Bonferroni adjustments ($\alpha=0.008$) indicated that the slope of V-I (0.246) was again significantly lower than other slopes (see Table IV). The slopes of VL-I (0.491) and V (0.464) were not significantly different; however, the slope of VL (0.646) was significantly higher than that of V and approached being significantly higher than that of VL-I ($p=0.015$).

Inspection of Fig. 4, which compares loudness and overall level as a function of CVR in stimuli with VL consonants, indicates that the significant quadratic component apparent in the overall level functions exerted no appreciable influence on the loudness functions. However, in Fig. 5 there is a hint of the modest quadratic component in the loudness function for the stimuli with V consonants (but not for the stimuli

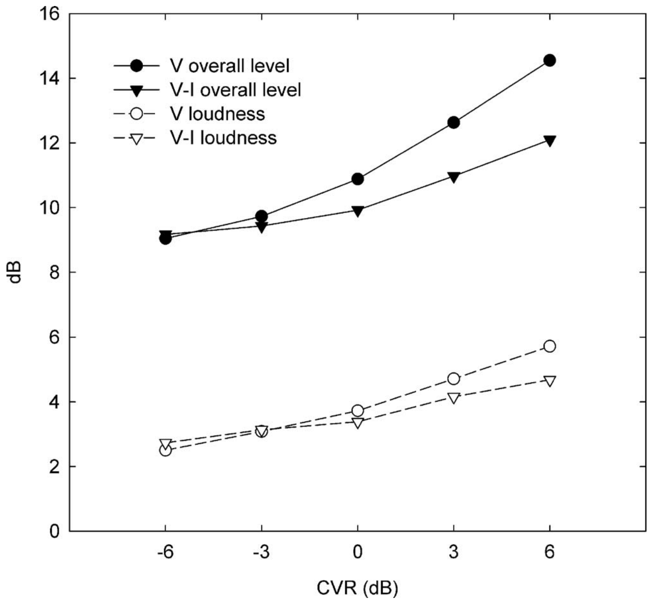

FIG. 5. Comparison of loudness and overall level (in arbitrary-reference dB values) as a function of consonant-vowel intensity ratio (CVR) for stimuli with both voiced (V) consonants and stimuli with only the initial voiced consonant (V-I) altered. Functions relating loudness and overall level to CVR are separated arbitrarily for ease of viewing.

where only the initial consonant was varied) when these curves are compared to the corresponding quadratic overall level curves.

Another way of examining the loudness functions is to look at differences in the end points of the CVR functions. These are seen in Table VI, where the values at −6 dB CVR are subtracted from those at +6 dB, for each loudness and overall level function. The overall level increases range from 2.9 dB for V-I to 7.7 dB for VL, with corresponding linear slopes of 0.246 and 0.646 (from Table IV). Similar calculations for loudness are also shown. The loudness increase ranges from 2.0 dB for V-I to 3.6 dB for VL, with corresponding linear slopes of 0.164 and 0.306. Note that the net increase in loudness for the four word types is roughly half the net increase in overall level, with the possible exception of V-I.

Table VII shows the slopes for loudness and overall level for individual stimulus words having both consonants modified and only initial consonants modified. It can be seen that the overall level slopes for the stimuli in the initial-only conditions are lower than the corresponding slopes when both consonants were modified. This is true for every individual word in both the voiced and voiceless conditions and is reflected in the group means. This is to be expected because a smaller portion of the word was modified in the initial-only conditions. The loudness slopes show a similar pattern. Again, this is true for every individual word in both conditions and is reflected in group means.[4] These patterns suggest an influence of overall level on the loudness slopes. For example, the lowest loudness slope (0.164) is associated with the lowest overall level slope (0.246). Thus, overall level had some influence on the loudness functions. However, the relative amounts of influence between vowel dominance (slope=0.0) and overall level (loudness slope

TABLE VI. Net increase in dB from −6 to +6 dB CVR conditions for overall level and loudness for all four word types, as seen in Figs. 2 and 3.

	Overall level (dB)			Loudness match (dB)		
Word type	−6	+6	Net increase	−6	+6	Net increase
VL	3.8	11.5	7.7	2.5	6.1	3.6
VL-I	2.6	8.5	5.9	3.0	6.1	3.1
V	9.0	14.5	5.5	2.5	5.7	3.2
V-I	9.2	12.1	2.9	2.7	4.7	2.0

TABLE VII. Slopes of functions relating loudness to consonant-vowel intensity ratio (CVR) and slopes relating overall level to CVR. Shown are values for each word in each condition as well as condition means. Also shown are ratios between loudness slopes and overall level slopes, for each word, and those based on condition means.

Word	Loudness slope	Overall level slope	Ratio
	VL		
Chaff	0.33	0.571	0.578
Chef	0.367	0.691	0.531
Church	0.341	0.614	0.555
Fish	0.28	0.722	0.388
Fuss	0.299	0.666	0.449
Sash	0.281	0.589	0.477
Sauce	0.316	0.509	0.621
Sis	0.332	0.777	0.427
Surf	0.294	0.680	0.432
Thief	0.222	0.640	0.347
Mean	0.306	0.646	0.474
	VL-I		
Chef_i	0.317	0.536	0.591
Church_i	0.322	0.429	0.751
Fish_i	0.22	0.440	0.500
Sis_i	0.271	0.584	0.464
Thief_i	0.203	0.467	0.435
Mean	0.267	0.491	0.544
	V		
Kneel	0.168	0.349	0.481
Lamb	0.221	0.347	0.637
Lawn	0.149	0.347	0.429
Learn	0.308	0.466	0.661
Limb	0.382	0.544	0.702
Man	0.155	0.396	0.391
Moon	0.318	0.509	0.625
Numb	0.264	0.532	0.496
Room	0.447	0.588	0.760
Rum	0.274	0.557	0.492
Mean	0.269	0.464	0.580
	V-I		
Lawn_i	0.132	0.134	0.985
Man_i	0.086	0.173	0.497
Numb_i	0.142	0.285	0.498
Room_i	0.315	0.334	0.943
Rum_i	0.145	0.305	0.475
Mean	0.164	0.246	0.667

=overall level slope) remain to be examined. To quantify this relationship, ratios of the loudness slope to the overall level slope were calculated. That is, if the loudness slopes were very high (i.e., approaching that of the overall level), the ratio would be near 1.0. Similarly, a very small ratio approaching zero would indicate vowel dominance. Table VII shows that the calculated ratios based on group means are all very near 0.5, indicating similar influence of overall level and vowel level on CVC word loudness. These ratios based on group means characterize well the ratios for individual words, as values are within 0.15 of the mean in 22 of 25 cases. The exception is condition V-I, which has a mean ratio of 0.667, due to two words having similar loudness and overall level slopes.[5]

IV. DISCUSSION

The two sources of speech loudness under consideration in the present study are the overall level of the monosyllabic word and the level of the vowel. The fact that the level of the vowel was held constant and the level of the consonants (and thus the overall level) was varied allows the generation of two possible slopes for the loudness functions. If the vowel was the sole determinant of CVC syllable loudness, as suggested by Montgomery *et al.* (1987), the slope of the loudness function would be zero. That is, the consonants, at least in the range of −6 to +6 dB CVR, would show no influence on the loudness judgments. On the other hand, if the overall level of the word was the only factor in CVC loudness, the slope for each word type would be the same as the corresponding slope of the function relating overall level to CVR, including the nonlinear component seen at the higher CVRs. Figures 4 and 5 show this not to be the case. The slopes of the functions relating loudness to CVR for all word types are greater than zero (see Table IV), indicating consonant influence.

There are several possible sources for the disagreement between the present study and that by Montgomery *et al.* (1987), which indicated vowel dominance. First, the very high presentation level (90 dB SPL) employed in the earlier study undoubtedly elicited the acoustic reflex, which becomes prominent at levels over 80 dB SPL (Durrant and Lovrinic, 1995, p. 172; Pickles, 2008, p. 23). The acoustic reflex is compressive in function and acts to reduce the loudness of sounds. Thus it may have reduced the impact of the amplified consonants in Montgomery *et al.* This interpretation is consistent with the fact that the listeners in Montgomery *et al.* also judged non-speech stimuli with amplitude contours similar to the CVC words and produced a slope of only 0.1. Another difference potentially arises from the unusual mixture of voiced and voiceless initial and final consonants in the Montgomery *et al.* study, including stops that were only altered in one of the two consonant positions in each word. The amplification of stop consonants may be expected to have only minor influences on loudness, as much of the signal is silent. Also, the current data indicate that modification of one consonant in a CVC word (relative to both consonants) results in reduced influence of consonant intensity on loudness. Finally, Montgomery *et al.* used the single word "laugh" as the standard for all CVC loudness judgments, thus requiring listeners to judge loudness across different vowels and consonants. In the present study, a more direct measure of CVR was obtained, by pairing only the unaltered form of a word as the standard for that word. Thus "sis" in unaltered form was always compared to "sis" with altered CVRs.

Figure 2 shows that the overall level of the stimuli increased monotonically with increasing CVR, as would be expected, with a modest quadratic component in a log-log plot. The curves for V, VL, and VL-I are generally parallel; the V-I stimuli, however, showed less of an effect of increasing CVR. This is apparent in Table VI, where the overall level of the word at +6 dB CVR is subtracted from the overall level at −6 dB CVR. The increases range from 2.9 dB for V-I to 7.7 dB for VL, with corresponding linear slopes of 0.246 and 0.646 (from Table IV). Thus it appears that listeners use overall level to some extent in their loudness judgments. This conclusion is supported by the mean ratios of the slopes of loudness to overall level, which show approximately equal influence of overall level and vowel level, ranging from 0.474 to 0.667 (Table VII).

The voiced stimuli with only the initial consonant varied (V-I) yielded the lowest slope of the loudness function (0.164), which would appear to indicate more dependence on the vowel than the overall level. However, the slope of the overall level function was also extremely low (0.246). The overall level function was flatter than that obtained when both consonants were modified, because the voiced stimuli were characterized by initial consonants that were very short in duration, which exerted reduced influences on overall level. The slope of the overall level dropped from 0.464 to 0.246 when the final consonant was not amplified—an approximate 50% reduction. However the effect of not amplifying the final consonant in the stimuli with voiceless consonants showed only a 25% reduction in slope of overall level because of the influence of the longer voiceless consonants.

Another question concerns the contribution of the initial consonant to speech loudness. In two experimental conditions, only the CVR of the initial consonant was altered, leaving the final consonant in its lower, unaltered state. Obviously if neither initial nor final consonant were altered and the vowel was unaltered, the slope of the loudness function across CVRs would be close to zero. Thus the increase from 0.0 to 0.164 for the stimuli with V-I consonants and the increase from 0.0 to 0.267 for the stimuli with VL-I consonants represents the (statistically significant) contribution of the initial consonant. These numbers across V and VL conditions were significantly different, indicating more influence of altering the longer voiceless initial consonants than voiced initial consonants. Similarly, when the final consonant was altered in addition to the initial consonant, the slopes increased significantly from 0.164 to 0.269 for words with voiced consonants (V-I, V) but non-significantly from 0.267 to 0.306 for words with voiceless consonants (VL-I, VL). Thus, the contribution to loudness of the initial V consonant (seen in the significant difference from zero of the V-I slopes) and final V consonants (seen in the difference from

V-I to V) seems to be approximately equal. However, the VL consonants only showed a significant effect of the initial consonant. The rationale for including both initial-consonant conditions (VL-I and V-I) and both-consonant conditions (VL and V) was to provide different ways to manipulate overall level while keeping the vowel level constant. This allowed the influence of overall level on loudness to be examined in detail. Given the current results, however, it might be of interest to confirm the different contributions of initial verses final consonants on word loudness by also manipulating only the final consonants.

As seen in Table VII, the loudness slopes for the stimuli with both V and VL consonants modified are quite similar (0.269 and 0.306, respectively). Recall that the early discussions of vocal effort identified SGP during vowel production as a source of effort that can potentially influence speech loudness judgments. However, it is likely that effort as perceived by the talker and later associated with loudness is composed of at least two components for voiced sounds: the true SGP sensed in the trachea and the muscular tension necessary to produce and maintain vocal fold approximation (Allen, 1971). For voiceless consonants, the glottis is open and the intraoral pressure behind the lingual or labial constriction is the same as in the trachea. Arkebauer *et al.* (1967) showed that the intraoral air pressure during the production of /s/, for example, was 5.8 cm H_2O, very similar to SGP during normal vowel and sonorant production, which averages approximately 7 cm H_2O (Behrman, 2007, p. 164). Thus, words with voiced consonants and words with voiceless consonants have similar air pressures in the trachea. The fact that the loudness slopes of the stimuli with V and VL consonants are similar suggests that the SGP necessary for voicing during the initial and final consonants may play a role in loudness judgments that is similar to the tracheal and intraoral air pressure during voiceless consonant production. (Although, as noted above, this influence on loudness is shared with the overall level of the monosyllabic word.) The original study by Lehiste and Peterson (1959), supporting the influence of effort on speech loudness, used isolated vowels or held the consonant context constant and thus was not able to separate the effects of articulatory mechanics, including intraoral pressure during voiced and voiceless sound production. In the present study, the variation of CVR allows the interpretation that one source of judgment in the loudness of monosyllabic words is general articulatory effort, including effects of articulation on intraoral as well as tracheal pressure, in addition to the overall intensity of the word.

In addition to overall level and articulatory effort, three other possible influences on the loudness slopes must be considered. One involves the temporal integration properties of the auditory system. It is well known that a short duration signal has a higher absolute threshold and may be perceived as less loud than a longer stimulus of equal overall level (e.g., Nishinuma *et al.*, 1983; Florentine *et al.*, 1996). While effects on absolute threshold vary within a range from 1 ms to approximately 250 ms, effects on loudness can be observed across longer durations. This is relevant to the present study because the words containing voiced consonants had shorter consonants, longer vowels, and shorter overall durations than words containing voiceless consonants. However, it is first important to note that effects on suprathreshold loudness are typically observed for stimuli having far larger duration differences than those employed here. Perhaps more importantly, the loudness of each altered word in the current study was judged only against the loudness of that same word in unaltered form, eliminating any temporal difference across loudness matches. Thus, even if different amounts of temporal integration had occurred across different words, the difference in CVR was the only variable contributing to differences in loudness, because the duration of the standard and the stimulus was identical. A potentially desirable follow-up experiment could include stimuli in which vowel and consonant duration is examined in greater detail.

The second possible influence on the loudness slopes is the effect of increasing bandwidth that occurs when consonants, especially voiceless consonants, are amplified. The amplification (although linear) results in a wider audible bandwidth in the higher frequencies, due to the shape of the long-term spectrum of speech, which may affect the loudness slopes. This increase in bandwidth could have the effect of increasing loudness, especially in the +3 and +6 dB CVR alterations and may explain some of the slight increase in slope for the stimuli with VL consonants. But again, the difference between the loudness slopes derived from stimuli with V and VL consonants was small and nonsignificant. Therefore, the effects of increased bandwidth may be present, but did not appear to play a major role in the observed judgments.

The final consideration involves nonsimultaneous (forward and backward) masking. The possibility exists that the more intense consonants might have partially masked the vowels at higher CVRs, thus affecting loudness judgments. Backward masking, the weaker of the two phenomena, was likely not a factor because it diminishes rapidly after approximately 25 ms (Wilson and Carhart, 1971) and is often eliminated with practice (Oxenham and Moore, 1994). Although forward masking remains a possibility, the masker must be far more intense than the signal for the effect to occur. Consider the forward masking of a click by broadband noise (more analogous to speech-on-speech masking than the more common tone-on-tone masking). Wilson and Carhart showed that a 70 dB SPL noise elevated click detection 13 dB above quiet threshold at a delay of 40 ms. That is, if the unmasked threshold of the click was 20 dB SPL, then it was detected reliably at 33 dB in the presence of the 70 dB masker, yielding an SNR of −37 dB. Thus, at a CVR of +6 dB in the present study, forward masking would likely be quite limited and not substantially affect the loudness of the vowel or word.

In conclusion, varying the level of the initial consonant and of the initial and final consonants in CVC words indicates that the overall level of the word and the level of the vowel contribute approximately equally to the loudness of monosyllabic words. Thus, speech loudness appears to be based to some extent on features unique to speech, such as sensations experienced during speech production. The current results should serve to further our understanding of the basic processes by which listeners judge the loudness of speech. It also may have more practical applications. Hearing aids use compression algorithms to fit the wide dynamic range of acoustic speech energy into the often smaller dynamic range of residual hearing. Cochlear implants also employ compression to fit the range of acoustic energy into the far smaller neural dynamic range. A better understanding of the determinants of speech loudness may be of value in the design of algorithms that provide more natural or comfortable loudness percepts.

ACKNOWLEDGMENTS

The current work was submitted by the first author in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of South Carolina. It was supported in part by NIH Grant No. DC008594 to EWH. MATLAB programming by Philippe Gresle and manuscript preparation assistance by Carla Berg are gratefully acknowledged.

[1]The vowel in words, "church," "surf," and "learn" is /ɝ/, and does not contain the postvocalic liquid consonant /r/.

[2]The intensity measurement of the final affricate in the word "church" included the silent closure gap because the gap is a key element in the recognition of final affricates. The level without the silent gap was found to be within 1.04 dB of the level with the gap, which aided in the decision to use the level from the longer section.

[3]Additional analyses were conducted to examine potential differences in loudness matches produced by the older versus younger subjects. First, the absolute difference between the mean loudness slopes produced by the two older versus the eight younger subjects, across the four word types, was found to be 0.03, with no difference for a word type greater than 0.05. Second, the mean slopes for the 30 individual words were found to be highly correlated across the two subject groups ($r=0.91$, $p<0.001$). Finally, the raw loudness values (for 5 CVRs $\times$ 30 words $=$ 150 loudness matches) were found to be highly correlated across subject groups ($r=0.92$, $p<0.001$). It is concluded that the older subjects produced scores similar to those produced by the younger subjects.

[4]The large number of judgments collected in the current study necessitated that a limited number of words be employed. Although the trends are reasonably consistent, it should be kept in mind that the current conclusions are based on this limited number of stimulus words.

[5]It is possible to speculate that the high loudness/overall level slope ratios for the V-I forms of "room" and "lawn" (Table VII) are attributable to vowel rounding, which may have had the effect of increasing the influence of overall level on loudness, especially when only these initial consonants were modified. Indeed, the average slope ratios for words having rounded vowels were higher in every condition (considerably higher in the VL-I and V-I conditions) than the ratios for words having unrounded vowels.

Allen, G. D. (**1971**). "Acoustic level and vocal effort as cues for the loudness of speech," J. Acoust. Soc. Am. **49**, 1831–1841.

Arkebauer, H. J., Hixon, T. J., and Hardy, J. C. (**1967**). "Peak intraoral air pressures during speech," J. Speech Hear. Res. **10**, 196–208.

Behrman, A. (**2007**). *Speech and Voice Science* (Plural, San Diego), p. 164.

Blood, G. (**1981**). "The interactions of amplitude and phonetic quality in esophageal speech," J. Speech Hear. Res. **24**, 308–312.

Brandt, J. F., Ruder, K. F., and Shipp, T., Jr. (**1969**). "Vocal loudness and effort in continuous speech," J. Acoust. Soc. Am. **46**, 1543–1548.

Durrant, J. D., and Lovrinic, J. H. (**1995**). *Bases of Hearing Science*, 3rd ed. (Williams & Wilkins, Baltimore), p. 172.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Ladefoged, P., and McKinney, N. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech," J. Acoust. Soc. Am. **35**, 454–460.

Lehiste, I., and Peterson, G. E. (**1959**). "Vowel amplitude and phonemic stress in American English," J. Acoust. Soc. Am. **31**, 428–435.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Luce, P. A., and Charles-Luce, J. (**1985**). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," J. Acoust. Soc. Am. **78**, 1949–1957.

Mendel, M. I., Sussman, H. M., Merson, R. M., Naeser, M. A., and Minifie, F. D. (**1969**). "Loudness judgments of speech and nonspeech stimuli," J. Acoust. Soc. Am. **46**, 1556–1561.

Montgomery, A. A., Prosek, R. A., Walden, B. E., and Cord, M. T. (**1987**). "The effects of increasing consonant/vowel intensity ratio on speech loudness," J. Rehabil. Res. Dev. **24**, 221–228.

Nishinuma, Y., Di Cristo, A., and Espesser, R. (**1983**). "Loudness as a function of vowel duration in CV syllables," Speech Commun. **2**, 167–169.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Pickles, J. O. (**2008**). *An Introduction to the Physiology of Hearing*, 3rd ed. (Emerald Group, Bingley, England), p. 23.

Rogers, C. F., Healy, E. W., and Montgomery, A. A. (**2006**). "Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions," J. Acoust. Soc. Am. **119**, 2276–2287.

Silva, I., and Florentine, M. (**2006**). "Effect of adaptive psychophysical procedure on loudness matches," J. Acoust. Soc. Am. **120**, 2124–2131.

Stevens, S. S. (**1955**). "The measurement of loudness," J. Acoust. Soc. Am. **27**, 815–829.

Stevens, S. S. (**1956**). "The direct estimation of sensory magnitudes-loudness," Am. J. Psychol. **69**, 1–25.

Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**, 847–863.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: Interactions and additivity," J. Acoust. Soc. Am. **49**, 1254–1263.