On enhancement of spectral contrast in speech for hearing-impaired listeners

H. Timothy Bunnell

Citation: The Journal of the Acoustical Society of America **88**, 2546 (1990); doi: 10.1121/1.399976

View online: https://doi.org/10.1121/1.399976

View Table of Contents: http://asa.scitation.org/toc/jas/88/6

Published by the Acoustical Society of America

On enhancement of spectral contrast in speech for hearing-impaired listeners

H. Timothy Bunnell
Applied Science and Engineering Laboratories, Alfred I. duPont Institute, Wilmington, Delaware 19899

(Received 29 June 1989; revised 7 May 1990; accepted 21 August 1990)

A digital processing method is described for altering spectral contrast (the difference in amplitude between spectral peaks and valleys) in natural utterances. Speech processed with programs implementing the contrast alteration procedure was presented to listeners with moderate to severe sensorineural hearing loss. The task was a three alternative (/b/,/d/, or /g/) stop consonant identification task for consonants at a fixed location in short nonsense utterances. Overall, tokens with enhanced contrast showed moderate gains in percentage correct stop consonant identification when compared to unaltered tokens. Conversely, reducing spectral contrast generally reduced percent correct stop consonant identification. Contrast alteration effects were inconsistent for utterances containing /d/. The observed contrast effects also interacted with token intelligibility.

PACS numbers: 43.72.Ew, 43.71.Ky, 43.66.Ts

INTRODUCTION

Several recent studies have examined the enhancement of spectral peaks in speech and speech-like stimuli as a means of enhancing speech for hearing-impaired listeners (e.g., Boers, 1980; Summerfield *et al.*, 1985; Bustamante and Braida, 1986, 1987). In these studies, spectral contrast (roughly the difference in amplitude between peaks and valleys of the speech spectrum) is exaggerated to produce a spectrum in which a greater than normal proportion of the total spectrum energy is more narrowly focused around important spectral features such as formant frequencies. Expectations of improved recognition of contrast-enhanced speech by hearing-impaired listeners are based on the notion that increasing spectral contrast will partly compensate for the poorer than normal frequency resolution often accompanying a hearing deficit (Tyler *et al.*, 1980). Poor frequency resolution could result in a "blurred" internal representation of spectral information in which features of low spectral contrast would be lost. In this case, boosting spectral contrast could restore the discriminability or detectability of such features.

However, experimental studies using contrast-enhanced speech have demonstrated only marginal improvements in synthetic syllables and no improvements with natural speech. Boers (1980) exaggerated spectrum contrast in sentence length utterances using a procedure based on squaring spectrum levels and normalizing amplitude. High-amplitude regions of the spectrum will grow more in amplitude when squared than will low-amplitude regions. Several sentences from the speech reception test (SRT, Plomp and Mimpen, 1979) were processed and mixed with noise shaped to the long-term speech spectrum. The noise had also been processed and then mixed with the speech before presentation to normal-hearing and hearing-impaired listeners. For both listener populations, increased contrast resulted in poorer SRT scores. However, Boers did note that the contrast alterations introduced some distortion into the speech and it is possible that the distortion reduced utterance intelligibility more than enhanced contrast improved intelligibility.

Summerfield *et al.* (1985) varied spectral contrast by varying formant bandwidth in the synthesis of CVC syllables. Decreasing formant bandwidth for the synthesis increases the amplitude of formant peaks relative to surrounding spectrum levels (hence, increases contrast) and increasing formant bandwidth has the opposite effect of reducing the amplitude differences between formant peaks and the valleys between formants. A continuum of syllables varying in contrast (formant bandwidth) was synthesized; the excitation was a noise source to avoid problems with the coincidence of formants and harmonics. The range of bandwidths was chosen to include the bandwidths found in natural productions. Both normal hearing and hearing-impaired listeners were presented these stimuli in randomized lists for identification. Overall, identification improved as contrast increased from the lowest values to "normal" (i.e., stimuli with normal formant bandwidth), however, greater than normal contrast did not typically result in further improvement: only syllable-final consonants showed reliable gains in identification accuracy at heightened contrast values.

Bustamante and Braida (1986) reported a preliminary study that compared the effects of linear amplification, wideband compression, and contrast enhancement plus wideband compression. Their contrast enhancement technique was based on a principal components decomposition of short-term spectra. Contrast was enhanced by inflating the amplitude of higher-order principal components that are most strongly associated with narrow-band features of spectral shape. Four hearing-impaired listeners participated in two listening tasks, a CVC identification task and a sentence task in which correct keyword recognition was scored. The combination of wideband compression plus contrast enhancement did not generally result in better performance than wideband compression alone, and in one instance appeared to produce considerably poorer performance than either linear amplification or wideband compression.

While none of the reported results is especially encouraging, several factors warrant further evaluation of contrast enhancement techniques. For example, the Summerfield *et al.* (1985) results show a systematic relation between spectrum contrast and stop consonant identification over part of the range tested. This range limitation may apply more to synthetic than to natural speech since the latter is acoustically richer and more diverse than synthetic speech, and may yield different effects under contrast enhancement. Bustamante and Braida (1986) cited several qualifications to their findings including a possibly suboptimal choice of analyzing bandwidth, use of a constant expansion factor rather than one that, e.g., enhances narrower peaks to a greater extent, and the possibility that sharpening should not be applied to spectrally diffuse features. Boers (1980), as mentioned above, cited acoustic distortion as a mitigating factor in his results. The present study applied a low-distortion form of contrast enhancement, to natural speech. The enhancement technique used a dynamic filtering process to alter spectral contrast. The process was similar in its spectral consequences to that reported by Bustamante and Braida (1986). However, because of differences in implementation, the contrast enhancement could be applied differentially to portions of the speech spectrum, and probably produced sharper spectral features. The filtering scheme, described below, was implemented in software on a general purpose computer.

Before turning to a description of the filtering scheme and perception experiment, one other factor that motivated some of the acoustical analyses of the present study should be considered. A difficulty with the interpretation of results in studies that manipulate spectral contrast is the confounding of amplitude and contrast. Since contrast enhancement alters the amplitude of spectral features, positive results from contrast enhancement could be due to changes in amplitude, increased contrast, or some combination of both effects. This is an issue of particular interest in the case of hearing-impaired listeners. While it is clear that hearing impairment involves both an attenuation and a distortion component, the relative importance of each of these components remains unclear. In some cases, attenuation alone seems to account for the hearing deficit (e.g., Zurek and Delhorne, 1987), while in other cases some additional factor is needed to account for the effective hearing deficit (e.g., Walden *et al.*, 1981; Turner and Robb, 1987). If contrast enhancement results in improved speech reception, it is thus of interest to know the extent to which improvements can be attributed solely to increases in amplitude. In the present study, this issue is addressed via post hoc acoustic and statistical analyses. In particular, regression analyses are used to assess the independent contributions of both amplitude and contrast to perceptual effects associated with contrast alterations.

I. PERCEPTION EXPERIMENT

A. Method

1. Signal processing

The general procedure is to derive a filter that, when applied to speech will produce the desired alteration in spectral contrast. Since speech is generally nonstationary, the filter must also be nonstationary, adapting to changes in the speech spectrum over time. The process consists of two components: one is a standard digital filtering component, implemented via the fast Fourier transform (FFT), that filters the speech signal by advancing in small discrete steps through the digitized signal; the other component computes, for each time step, a new filter transfer function to be applied to the speech.

The FFT filtering component used 25.6-ms windows and 12.8-ms steps, thus a 50% overlap in adjacent analysis frames. With speech digitized at 10 kHz, this meant that each analysis window was 256 samples long and that adjacent analysis frames overlapped by 128 samples. On each step, the input sequence to be filtered was windowed (25.6-ms Hamming window) and transformed to a complex spectrum. Once the filter coefficients were computed (see below) they were applied to the complex spectrum and an inverse transform was performed to derive a filtered time signal. The first 128 samples of this sequence were added to the last 128 samples from the previous analysis frame and became the output sequence. The second 128 samples of the newly filtered time signal were then saved to be added into the overlapping part of the next analysis frame and so on. This procedure disregards wrap-around effects in the data that are due to the filtering, however, because the effects of the Hamming window were not removed prior to summing the overlapping sections of data, wrap-around effects did not produce perceptible waveform discontinuities.

The filter was derived by estimating the envelope of the log magnitude of the speech spectrum and from that deriving a target envelope containing the desired alterations in contrast. Each discrete frequency bin in the target envelope was computed using the following relation:

$$T_i = C \times (S_i - \overline{S}) + \overline{S},$$

where T_i is the target envelope amplitude at frequency bin i, S_i is the original envelope amplitude at frequency bin i, $\overline{S}$ is the average spectrum level, and C is a contrast alteration factor (hereafter contrast weight). Note that all spectrum levels are in decibels. When the contrast weight C is 1.0, the target envelope is the same as the original envelope. Contrast weights less than 1.0 produce contrast reduction while contrast weights greater than 1.0 produce contrast enhancement. In the extreme case of reduction ($C = 0.0$), all features of the original envelope would be lost in the target envelope; it would be flat at the average spectrum level. A contrast weight of 2.0 was used for the enhanced stimuli in the present experiment, and a weight of 0.0 was used for the contrast-reduced stimuli.

Once the target envelope was obtained, filter weights were computed by differencing the new and old log magnitude envelopes and converting to linear coefficients. This provided the frequency response of the filter that was used to map an input spectrum onto the desired output spectrum. The filter was then applied to the complex DFT of the original frame of speech. Thus the speech output for each analysis frame was a filtered version of the input with each component of the original signal potentially altered in amplitude, but unaltered in phase.

One additional aspect of the signal processing used in stimulus generation was a nonuniform contrast weight versus frequency profile. This allowed contrast to be altered primarily in the midfrequency portion of the spectrum while leaving both low- and high-frequency regions less affected. This was introduced because, on the average, energy in the speech signal is heavily weighted to the low frequencies and it would have been undesirable to allow further amplification of low-frequency peaks. Amplification of strong low-frequency components would have had the effect of saturating the dynamic range of the 12-bit signal digitization and could also have resulted in increased masking of higher frequency peaks by low-frequency peaks (Danaher and Pickett, 1975). Further, because of the particular importance of information in the region of $F2$ and $F3$ for perception of place of articulation, it was thought most valuable to concentrate on that spectral region.

The differential weighting of contrast was accomplished by scaling the filter coefficients so that the full effect of the contrast altering filter would apply only at the center of the spectrum (2.5 kHz) and was tapered to zero effect at both the lowest and highest frequencies. A Hamming window function applied to the log differences between target and input spectrum envelopes was used for this purpose. That is, the windowing function was applied over the linear frequency range from 0 to 5.0 kHz with the center of the window at 2.5 kHz giving unity gain to the effects of contrast alteration at that frequency and with the tails of the window tapering the magnitude of the alteration to 0.0 at the upper and lower limits of the frequency range. Figure 1 shows spectra from the midvowel region of one of the experimental stimuli before [panel (a)] and after [panel (b)] application of contrast enhancement. Note the relatively greater amplitude of the spectrum in the $F2$–$F3$ region and the compensatory decrease in amplitude throughout other parts of the spectrum.

2. Subjects

Ten sensorineural hearing-impaired students, undergraduates at Gallaudet University, served as paid listeners in this study. The mean audiogram in dB SPL for this group, obtained via a Békésy audiometer is shown in Fig. 2 as the upper curve (circles). The error bars show the plus and minus 1 standard deviation (s.d.) range at each of the audiometric frequencies. All but one of these listeners would be characterized as having a sloping loss. The listeners were part of a pool of paid listeners each of whom participated in from 2 to 4 sessions per week throughout the course of a semester. At the time when this study was run (midsemester), all listeners had extensive practice on the experimental task (stop consonant identification) and had participated in other studies using the original tokens of this study.

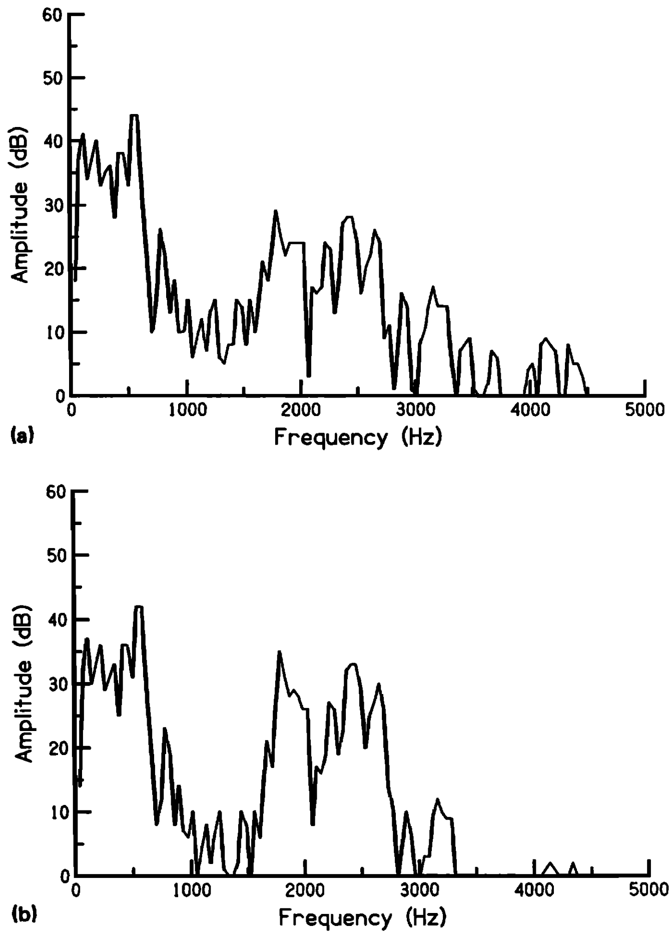

FIG. 1. Original (top) and contrast-enhanced (bottom) spectra. The enhancement was applied differentially to midfrequency peaks.

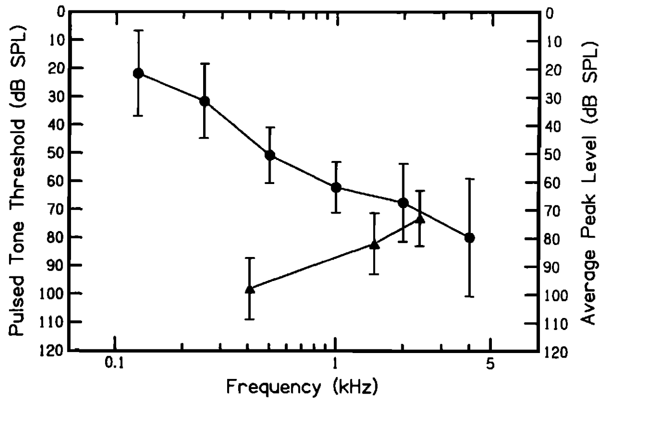

FIG. 2. Average pulsed-tone thresholds in dB SPL for the ten subjects obtained via Békésy audiometer (closed circles) and average amplitudes of spectral peaks associated with the first three formants of the experimental stimuli plotted at their average frequencies (closed triangles). Error bars indicate plus/minus 1 s.d. for both curves.

3. Stimuli

Three tokens of each utterance of the type "say a CVwuh" where $C = \{/b/,/d/,/g/\}$ and $V = \{/i/,/a/,/u/\}$ produced by a single male talker were selected from a larger corpus of utterances. These utterances had been ranked on the basis of the intelligibility of the stop consonants. The ranking had been obtained from studies in which the tokens were presented under noise masking to normal hearing listeners for identification (Bunnell and Martin, 1988). The three tokens chosen for this experiment were known from the ranking data to vary in intelligibility of their target stop consonants and will be referred to as GOOD, MODERATE, and POOR intelligibility. It should be noted that this intelligibility classification is relative to other instances of the same stop consonant in the same vowel context. For example, a "GOOD" /gi/ token is not necessarily more intelligible than a "MODERATE" /ba/ token because its rating is relative only to other instances of /gi/ in the original corpus.

From each of these three tokens, two additional tokens were derived by using the contrast altering program either to enhance or reduce the amount of spectral contrast in midfrequency regions. The processing for contrast alteration was applied throughout the entire stimulus duration including all carrier phrase context. For the contrast-reduced versions, the contrast weight was set to zero, while for the contrast-enhanced versions the weight was 2.0 (i.e., contrast was increased by two times the difference in dB between the original envelope and the average spectrum level). After each altered-contrast token was computed, its amplitude was adjusted to the maximum level possible without incurring digital clipping. This ensured that each token utilized the maximum dynamic range of the system. In all, there were 81 stimuli in the experiment: three levels of CONTRAST (normal, reduced, enhanced), by three levels of CLARITY (good, moderate, poor), by three STOP consonants by three VOWELS.

4. Procedure

Listeners were seated in an IAC booth facing a Fluke Infotouch CRT terminal with touch sensitive screen. Response alternatives for the identification task were displayed on the CRT screen and the listener was required to touch the alternative on the screen to register his/her response. Stimuli were presented to the listeners' audiometrically better ear through a TDH-39 earphone with MX41/AR cushion.

Stimuli were stored on a computer disk and presented in random sequence by a program that also monitored and recorded listener responses. Stimuli were blocked for presentation by vowel context, that is, stimuli presented in a given block of trials shared the same vowel, either /i/, /a/, or /u/. Each block of trials consisted of three repetitions of each of the 27 stimuli in the block, i.e., 81 trials. Correct answer feedback was not given. The listener was allowed up to 10 s to respond after the presentation of a test utterance before initiation of the next trial. After recording the listener's response, the system delayed for 750 ms and then proceeded to present the next trial stimulus. Thus the rate of stimulus presentation was largely paced by the listener's response time.

Listeners heard each of the blocks of stimuli a minimum of three times on different days. The order in which blocks were presented to each listener was rotated so that all listeners heard each vowel context in each ordinal position. On days when listeners were presented stimuli for this study, they were generally presented stimuli for other experiments as well.

Stimulus presentation level was based on each listener's most comfortable level (MCL) as determined from a procedure that adaptively varied the level of one of the test stimuli until a stable level was reached. The average MCL for the ten subjects in this experiment was 114 dB with a range from 95–130 dB. A programmable attenuator was used to adjust the playback level of each token to MCL. The rms amplitude of the waveform integrated within a 300-ms window centered on the consonant closure interval was used as the reference level in determining the attenuator setting. The lower curve (triangles) in Fig. 2 plots the average amplitudes (the decibel values were averaged) of spectral peaks measured at three discrete locations in all the experimental stimuli and scaled to a presentation level of 114 dB. These peaks were measured directly from digitally computed power spectra with no spectral smoothing or averaging and corresponded generally to very narrow-band (approximately 39-Hz bandwidth) estimates of the amplitudes of the first three formants. The average amplitudes are displayed at the mean frequencies of the raw peaks. The error bars show the 1 standard deviation range in the amplitude data. Since these data formed the basis for amplitude measures described below, a more complete description of how they were obtained may be found in the description of the methods for the acoustic analyses.

B. Results

For analysis of variance, data from the multiple sessions of each listener were combined to determine percentage of correct identifications for each token. Because the number of sessions differed across subjects, two analyses were run: one in which the averaged data for each subject were weighted in the ANOVA by the number of sessions for that subject, the other using the unweighted average data. The results of the two analyses were very similar and so the means and F ratios reported here are those of the unweighted analysis, however, no effect or interaction was considered significant if it did not appear as significant in both analyses.[1] The factors for this within-subjects design were CONTRAST (reduced, normal, enhanced), CLARITY (good, moderate, poor), STOP (b, d, g), and VOWEL (a, i, u). A summary table of all significant ($p < 0.05$) main effects and interactions is presented as Table I.

Figure 3 shows the mean percentage correct ID for each level of CONTRAST with percent correct ID increasing from reduced to normal and from normal to enhanced contrast. This effect was significant ($F[2,18] = 34.09, p < 0.01$)

TABLE I. Analysis of variance summary table for spectral contrast enhancement.

Source	*df*	*F*	*p*
Contrast	2,18	34.09	0.0000
Clarity	2,18	39.90	0.0000
Vowel	2,18	4.63	0.0239
Contrast × clarity	4,36	5.05	0.0025
Contrast × stop	4,36	5.72	0.0011
Contrast × vowel	4,36	4.54	0.0045
Stop × vowel	4,36	7.14	0.0002
Contrast × clarity × stop	8,72	3.88	0.0008
Contrast × stop × vowel	8,72	2.27	0.0316
Contrast × clarity × vowel	8,72	2.95	0.0065
Contrast × clarity × stop × vowel	16,144	2.92	0.0003

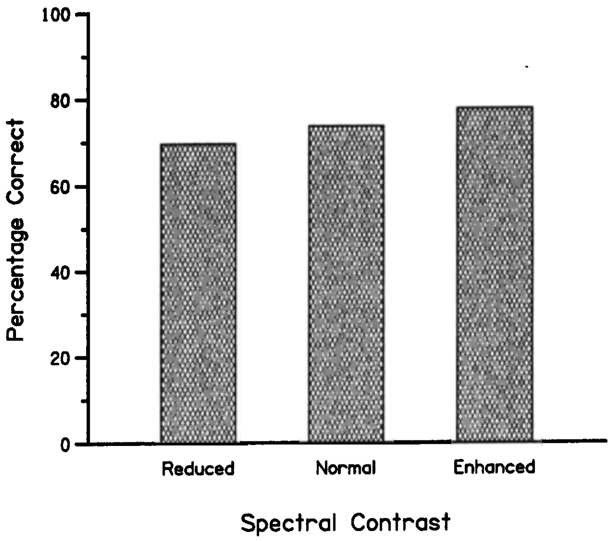

FIG. 3. Percentage correct stop consonant identification for each contrast condition.

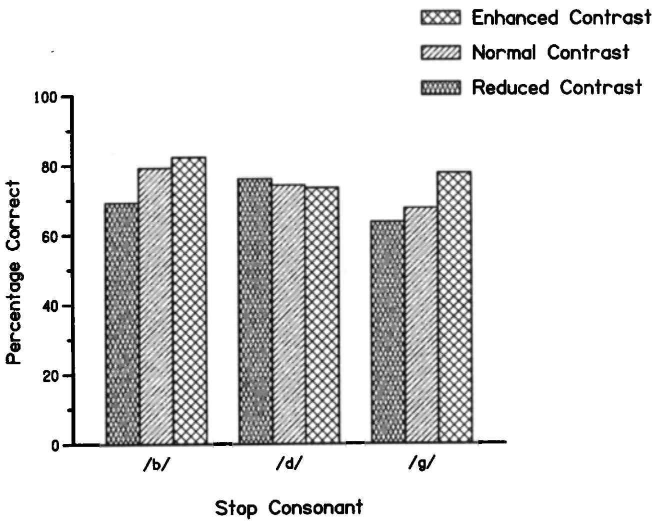

FIG. 5. Percentage correct stop consonant identification for each stop consonant in each contrast condition.

and all three of the pairwise differences between means in this figure are significant by Tukey HSD *post hoc* test. Figure 4 illustrates the significant interaction between the effects of CONTRAST and CLARITY ($F[4,36] = 5.05$, $p < 0.01$). Bars in this figure show the mean percentage correct ID for each of the three levels of CONTRAST separately for each level of token CLARITY. For the clearest tokens identification is uniformly good and increased contrast has little effect. The poorest tokens similarly show smaller effects from contrast variation. The greatest effect of contrast alterations was on tokens of intermediate clarity.

There were significant differences in the effects of contrast alterations on the different stop consonants as evidenced by the significant STOP by CONTRAST interaction ($F[4,36] = 5.72, p < 0.01$). Means from this interaction are shown in Fig. 5 and indicate that /b/ and /g/, but not /d/ showed improved identification following contrast enhancement and poorer identification when contrast was reduced.

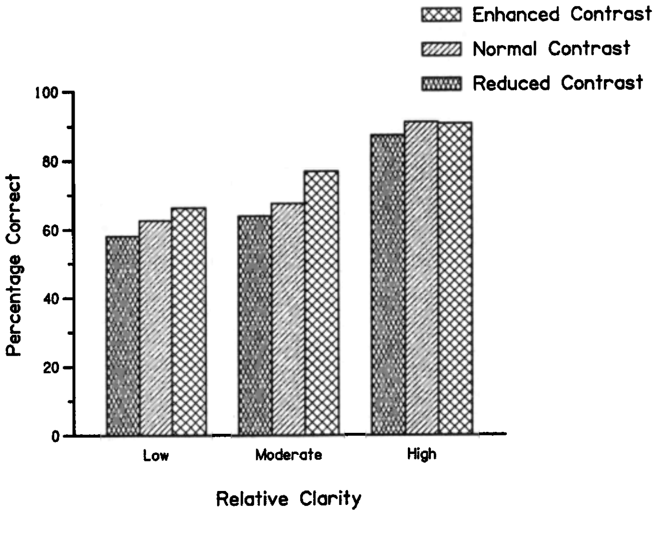

FIG. 4. Percentage correct stop consonant identification of low, moderate, and high intelligibility items for each contrast condition.

Many other interactions were significant, including the four way interaction between CONTRAST, INTELLIGIBILITY, STOP, and VOWEL ($F[16,144] = 2.92$, $p < 0.01$) which suggests that there were significant differences at the individual stimulus level.

C. Discussion

For listeners with moderate to severe sensorineural hearing loss, reduced midfrequency spectral contrast in short nonsense utterances led to poorer identification of intervocalic voiced stops. A similar result has been reported by Summerfield *et al.* (1985) for synthetic stop consonants, and by Dubno and Dorman (1987) in tests of synthetic vowel perception. In addition, enhanced midfrequency contrast led to improved identification of the naturally produced stops in our experiment. The overall significant improvement in stop consonant identification for contrast-enhanced tokens, however, was due to favorable effects on the perception of /b/ and /g/ tokens only: overall, contrast alterations had little effect on the correct identification of /d/ tokens.

In addition to limitations related to phonetic content, it would seem that not all tokens of a particular type are equally amenable to improvement through contrast enhancement. The best and worst exemplars of each phonetic category showed less improvement than did the moderate CLARITY exemplars. This may be due to two factors that clearly interact in speech perception: the information contained in the acoustic signal, and the accessibility of that information to the listener. If we assume that the effect of contrast enhancement is to make stimulus information more accessible to the listener, then it will work best in cases where stimulus information is unambiguous but inaccessible. Contrast enhancement alone is unlikely to be very helpful in the case that the stimulus information is, to begin with, ambiguous. That is, increasing the perceptual accessibility of phonetically ambiguous information is unlikely to reduce its phonetic ambiguity. Thus, in the present experiment, some tokens of poor CLARITY may have been ambiguous at the acoustic-phonetic level and not amenable to improvement via contrast enhancement. For the good CLARITY tokens the absence of improvement appears to be a simple ceiling effect.

It is also possible, of course, that the contrast altering program simply did not produce the desired acoustic effects on some stimuli. Consequently, acoustic measurements of all the stimuli were undertaken for use in determining the relationship between acoustic changes and the perceptual effects of those changes. For example, it is possible that overall weak effect of contrast changes on /d/ tokens stemmed from weak or inconsistent changes in contrast for those stimuli. A further question to be addressed by comparing results of acoustic analyses with perceptual results is the extent to which changes in the amplitudes of spectral peaks alone can account for perceptual effects. The results of the acoustic analyses were used in regression modeling of the perceptual results. These are presented in the next section.

II. ACOUSTIC ANALYSES

Two types of measurements were made: (a) amplitudes of spectral peaks (usually formants) in the region of the consonant release and vowel onset; and (b) two direct estimates of spectral contrast averaged within several time regions.

A. Method

1. Segmentation

Figure 6 illustrates criteria chosen for segmentation points. The stimulus shown in the top panel of this figure is the utterance containing the /ga/ token of good CLARITY, while the token in the bottom panel contains the /ga/ token of poor CLARITY. With the majority of the good CLARITY tokens the closure interval, release burst, friction, and voicing onset were clearly present, and their marking unambiguous. Voice onset was marked at the start of the first pitch period following closure release for which at least $F1$ and $F2$

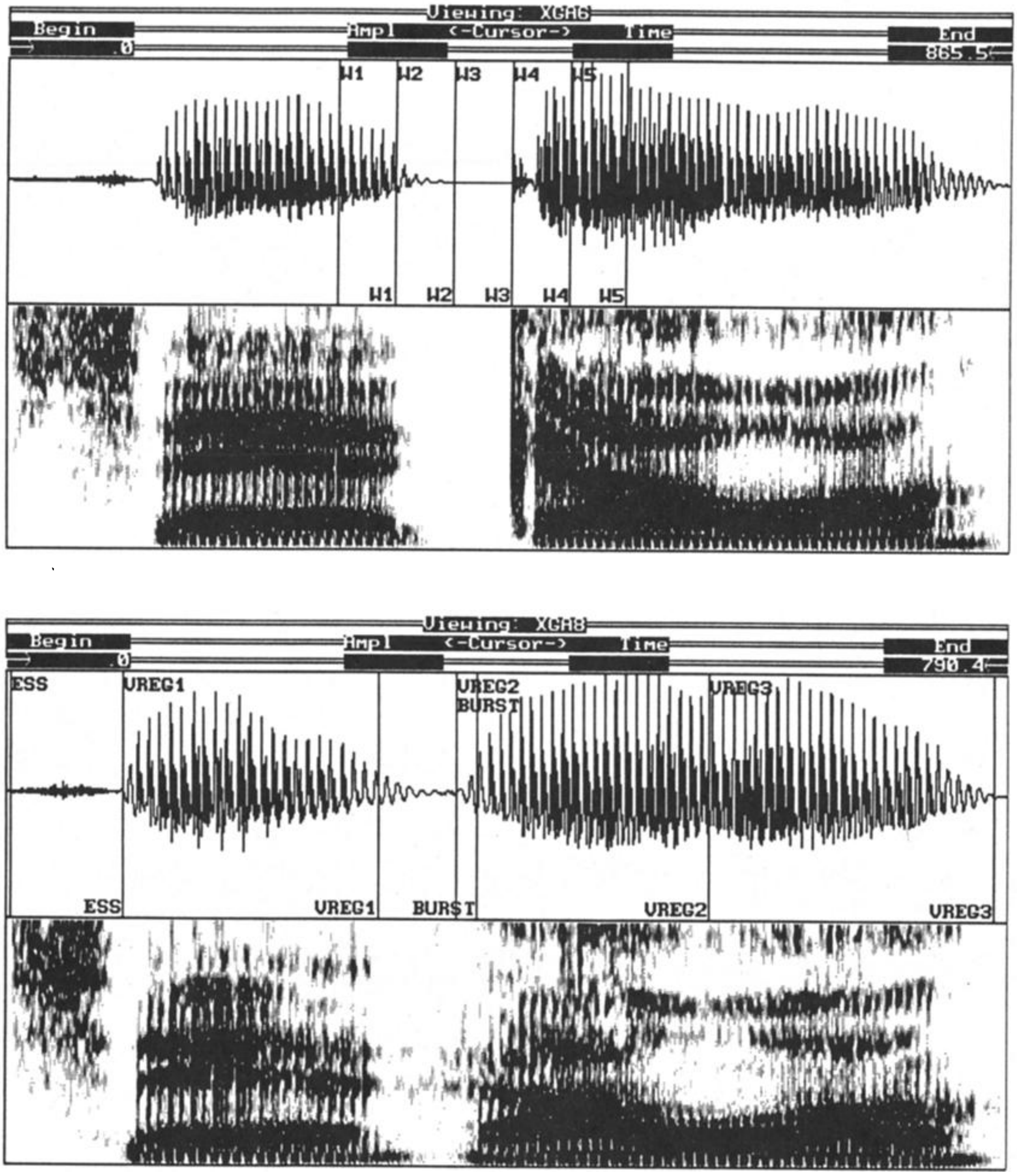

FIG. 6. Waveforms and spectra for representative GOOD (top) and POOR tokens. The total length of each utterance is shown under the word "End" on the upper right of each panel. Regions used in the acoustic analyses of the token are shown demarked by vertical lines. Labels at the top of a line indicate the beginning of the labeled region while labels at the bottom indicate the end of the region.

showed correlated pulse excitation in the spectrogram display. With tokens of poor CLARITY as illustrated in the bottom panel complete closure and silence was often not obtained, and in a number of cases no discernible release burst was present. For these tokens, the "burst" was marked as the start of the pitch period of lowest amplitude in the closure region and voice onset was marked as the start of the pitch period immediately following the burst.

2. Acoustic analysis

Two forms of acoustic analysis were used. First, based on segmentation marks for "burst" and "vowel onset," frequencies and amplitudes (in decibels) of five spectral peaks were measured at three points in time: at the location identified as the burst, at the location identified as the onset of voicing ($V0$), and at 50 ms after voicing onset location ($V50$). In most cases, the measured spectral peaks corresponded to the first five formants. Most of the exceptions to this were for burst segments within which one peak was always assigned to the highest magnitude peak below 1000 Hz. To ensure meaningful comparison of amplitudes between normal, contrast-enhanced, and contrast-reduced stimuli, corresponding peaks in each of the three versions were located based on their frequencies. In doing this, the peaks were first located in the contrast-enhanced versions and thereafter in the normal and contrast-reduced versions. The rationale for using the enhanced rather than normal tokens as the reference for peak location was the desire to choose locations most affected by the signal processing and hence most likely to relate to perceptual effects in the enhanced condition. However, it is doubtful that using the normal tokens as the reference would have resulted in the choice of different peaks.

Another analysis was used to obtain two different estimates of spectrum contrast. Contrast was defined as (a) the rms variation in the short term magnitude spectrum computed from a Hamming windowed 25.6-ms region of waveform, or (b) rms variation in the same spectrum after rescaling the frequency axis to Bark units (Zwicker, 1961) and smoothing to remove features of less than roughly two critical bandwidths. In both cases, rms values were computed as

$$rms = 10.0 \times \log \sum_{i=1}^{nbin} (S_i - \overline{S})^2,$$

where the summation is carried out over *nbin* frequency bins (in this case there were 128 frequency bins covering the range from 0–5 kHz), the S_i are the individual discrete spectrum levels and $\overline{S}$ is the average spectrum level for the frame. These data were computed for a series of frames at 10-ms intervals throughout the region surrounding the burst. Data for consecutive groups of five analysis frames were in turn averaged to provide average contrast estimates for each of the five temporal windows labeled $W1$ through $W5$ in Fig. 6.

3. Statistical analyses

Analyses of variance were used to identify reliable acoustic effects of the contrast alteration scheme. Separate analyses were run for the different types of acoustic measures (peak amplitudes and rms contrast) described above. In these statistical analyses, the 27 basic utterances were the sampling units so that CONTRAST was a "within sampling unit" effect while STOP and VOWEL were "between sampling unit" effects or grouping factors. For analysis of the peaks data, additional factors were measurement LOCATION (burst, $V0$, and $V50$), and spectral PEAK (1 through 5). For the rms contrast data, the two additional factors were the WINDOW [1 through 5 as per Fig. 5(top)], and contrast TYPE (Bark- or frequency-scaled). Differences in CLARITY were ignored in these ANOVA's so that there would be multiple samples per analysis cell.

Finally, all-subsets regression analyses were used to obtain models that predicted the perceptual effects of contrast alteration from measures of the acoustic effects. In particular, the dependent measure for each case was the difference in percentage correct identification between a contrast altered and unaltered version of each stimulus. The independent measures were the differences in the acoustic measures between the altered and unaltered versions. There were thus 54 cases for regression, 27 cases where a contrast-enhanced token was paired with its unaltered counterpart, and 27 cases where a contrast-reduced token was paired with its unaltered counterpart. There were in all 20 independent variables consisting of differences in contrast in each of the five windows (separate analyses were performed for each type of contrast measure), and differences in the amplitudes of each of the five spectral peaks at each of the three measurement locations.

This regression technique samples all possible models involving one or more of the independent variables, with the objective of finding a model that explains a sufficiently large proportion of the variance with a small number of terms. In choosing a model one is normally guided by a criterion measure that weighs variance explained against the number of terms in the model. This tends to favor the selection of models involving a smaller number of statistically independent terms over those involving a larger number of more interrelated variables. Once a model is chosen, statistics are available that estimate the significance of the overall fit and the significance of the contribution of the individual terms to the fit.

B. Results

1. Contrast measures

The signal processing produced significant changes in spectral contrast ($F[2,36] = 249.88, p < 0.01$) as expected, with average contrast increasing monotonically from the reduced to normal and normal to enhanced conditions. The two measures of contrast differed statistically, although not in very interesting ways: The frequency-scaled estimate of contrast was consistently larger in magnitude than the Bark-scaled and smoothed estimate ($F[1,18] = 714.41, p < 0.01$); and the TYPE of contrast estimate interacted significantly with the effect of enhancement ($F[2,36] = 438.89$, $p < 0.01$) reflecting the relatively greater differences in contrast when estimated without Bark smoothing. Both the main effect of contrast TYPE and the interaction are to be expected given the spectral smoothing associated with the bark-scaled data. While not significant, there was a trend toward smaller effects of contrast enhancement on /d/ tokens. All significant main effects and interactions for this ANOVA are listed in Table II.

2. Peak amplitudes

There were significant differences in the average amplitudes of the five peaks ($F[4,72] = 436.82, p < 0.01$) and in overall peak amplitude at each of the three measurement locations ($F[2,36] = 88.07, p < 0.01$). These effects reflect average spectrum slope and gross changes in amplitude from the closure region into the vowel. As would be expected, average peak amplitudes also interacted significantly with consonant ($F[8,72] = 4.83, p < 0.01$), with vowel context ($F[8,72] = 11.89, p < 0.01$), and with consonant and vowel context combined ($F[16,72] = 3.83, p < 0.01$).

The main effect of contrast alteration on average peak amplitude was small and not statistically significant ($F[2,18] = 2.05, p = 0.14$). Contrast enhancement resulted mainly in the removal of energy between spectral peaks. However, the interaction between contrast alteration and peak was significant ($F[8,144] = 68.81, p < 0.01$) with the effects of contrast alteration on the amplitude of peak 2 being generally opposite the effects seen for the other peaks. Means from this interaction are shown in Fig. 7. Here and throughout, the amplitude values are scaled per a 114-dB rms presentation level (i.e., the average MCL for the listeners in the perception experiment). Note that contrast enhancement produced some increase in the amplitude of peak 2 and tended to attenuate the other four peaks while contrast reduction had the reverse effect of attenuating peak 2 and amplifying other peaks. All significant terms of the analysis are shown in Table III.

Of particular interest are interactions that suggest differential effects of contrast alterations on the acoustics of the three stops. One such interaction involving consonant, contrast, position, and peak ($F[32,288] = 1.86, p < 0.01$), was the basis for the means displayed in Fig. 8. This figure shows, for each consonant, the average amplitude of each of the five peaks from the original tokens. Averages are shown for each of the three measurement locations (burst, $V0$, and $V50$). The largest differences in amplitudes were during the burst, and these differences appear to discriminate the stop conson-

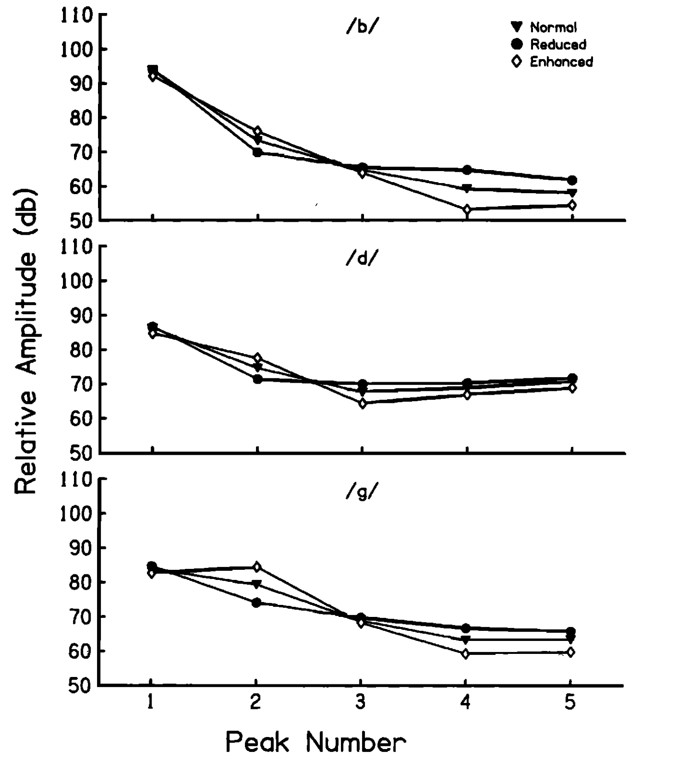

FIG. 7. Average peak amplitudes measured at the burst. Separate panels show data for each stop. Within each panel, the parameter is the contrast condition.

TABLE II. Summary of significant ANOVA effects for two types of contrast measurements.

Source	*df*	*F*	*p*
Vowel	2,18	14.50	0.0002
Contrast	2,36	249.88	0.0000
Window	4,72	59.27	0.0000
Type	1,18	714.41	0.0000
Vowel × contrast	4,36	3.30	0.0211
Stop × window	8,72	2.88	0.0077
Vowel × window	8,72	6.23	0.0000
Contrast × window	8,144	76.33	0.0000
Stop × type	2,18	5.85	0.0110
Vowel × type	2,18	8.01	0.0000
Contrast × type	2,36	438.89	0.0000
Window × type	4,72	56.28	0.0000
Stop × contrast × window	16,144	4.06	0.0000
Vowel × contrast × window	16,144	16.22	0.0000
Stop × vowel × type	4,18	6.58	0.0019
Stop × contrast × type	4,36	4.86	0.0031
Vowel × contrast × type	4,36	.68	0.0001
Stop × window × type	8,72	.37	0.0250
Vowel × window × type	8,72	18.10	0.0000
Contrast × window × type	8,144	5.49	0.0000
Stop × vowel × contrast × window	32,144	1.96	0.0041
Stop × vowel × contrast × type	8,36	5.18	0.0002

TABLE III. Summary of significant ANOVA effects for peak amplitude measurements.

Source	*df*	*F*	*p*
Location	2,36	88.07	0.0000
Peak	4,72	436.82	0.0000
Vowel × contrast	4,36	4.30	0.0000
Contrast × location	4,72	31.18	0.0000
Stop × peak	8,72	4.83	0.0001
Vowel × peak	8,72	11.89	0.0000
Contrast × peak	8,144	68.81	0.0000
Location × peak	8,144	24.22	0.0000
Stop × contrast × location	8,72	3.82	0.0009
Stop × vowel × peak	16,72	3.83	0.0000
Vowel × contrast × peak	16,144	4.64	0.0000
Stop × location × peak	16,144	4.14	0.0000
Vowel × location × peak	16,144	2.46	0.0026
Contrast × location × peak	6,288	4.58	0.0000
Stop × vowel × contrast × peak	32,144	4.14	0.0000
Stop × contrast × location × peak	32,288	1.86	0.0043
Stop × vowel × contrast × location × peak	64,288	1.47	0.0176

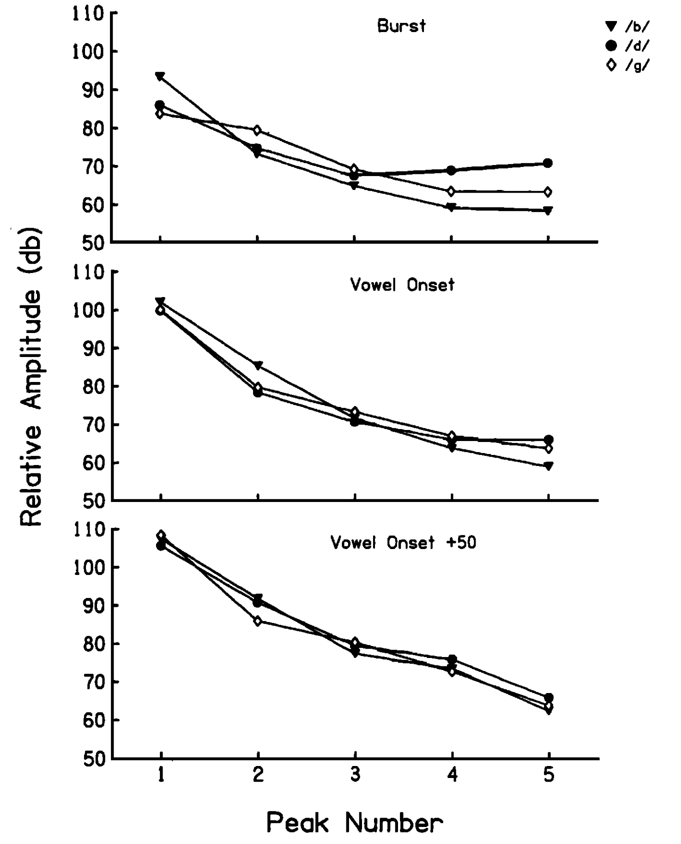

FIG. 8. Average peak amplitudes measured at each of three locations; Burst, vowel onset, and 50 ms after vowel onset. Separate curves are presented for each stop.

ants. The /b/ tokens were, on the average, stronger in amplitude than either /d/ or /g/ tokens at the lowest frequency peak, while /g/ tokens averaged the greatest amplitude of the three stops on the second and third peaks and /d/ tokens were strongest on the fourth and fifth peaks. This pattern is consistent with expectations for the gross shape of release spectra for these three stops (Blumstein and Stevens, 1979). There were additional significant effects from this analysis and those are summarized in Table II.

3. Regression modeling

Separate regression analyses were run for each of the two types of contrast estimates (frequency- and Bark-scaled). Variables in the regression were the contrast in each of the five windows, peak amplitude terms at the burst, at vowel onset, and 50 ms after vowel onset.

For the analyses using frequency-based contrast estimates, the variables comprising the "best" subset (Minimum Mallow's Cp criterion, Daniel and Wood, 1971) are listed in Table IV. The "variance contributed" column in this table reports the amount by which R^2 would be reduced if this term were removed from the regression equation. This model includes terms for the differences in amplitudes of peaks 1 and 2 at the burst, peak 2 at the vowel onset, and also 50 ms after vowel onset, and the difference in contrast in window 5. The test for the significance of individual terms showed that the term involving differences in the amplitude of peak 2 at the burst was not significant ($t = 1.53$). The term for amplitude differences in peak 2, 50 ms after vowel onset had the largest contribution to R^2 and had a *negative* regression coefficient. The raw correlation between this term and the dependent measure was small and positive ($r = 0.140$) suggesting that the importance of this term in the regression model was to explain residual variance in perceptual effects after variation related to other terms was removed.

TABLE IV. "Best" regression model using frequency-based contrast.

Name	Coef	t	p	R^2
Intercept	1.909	1.37	0.178	
Peak 1 (burst)	3.715	3.13	0.003	0.09213
Peak 2 (burst)	1.667	1.53	0.132	0.02204
Peak 2 ($V0$)	2.431	4.31	0.000	0.17415
Peak 2 ($V50$)	− 2.169	− 5.10	0.000	0.24421
Contrast ($W5$)	3.760	3.69	0.001	0.12748

Table V presents the "best" model obtained for Bark-based contrast variables. In this case, the term expressing differences in the amplitude of peak 2 at the burst was not present. There were no other noteworthy differences between the previous model and this one. The two models taken together suggest that four acoustic features were of particular importance in predicting perceptual effects of the contrast alterations. These were amplitude measures at consonant release and vowel onset, spectral contrast following vowel onset, and the amplitude of peak 2 (typically $F2$) following vowel onset. The latter feature appears to capture a trend in the data for high-amplitude $F2$ following vowel onset to weaken the otherwise positive relationship between peak amplitudes and ID scores. Such an effect could easily be seen if, for example, changes in peak amplitudes associated with contrast alternation were not proportional for both the closure interval and the vowel following closure. In that case, the vowel-to-consonant amplitude ratio would increase leading to a proportionately lower sensation level for information in the region of the burst.

III. GENERAL DISCUSSION

Based on the results of the perception experiment and acoustic analysis, two issues are of particular interest. First, is the question of whether any basis can be found for the inconsistent effects of contrast enhancement on /d/ in the present study. The second issue is the relative importance of formant amplitude (and presumably audibility) versus contrast in accounting for the present results.

TABLE V. "Best" regression model using Bark-based contrast.

Name	Coef	t	p	R^2
Intercept	1.275	0.94	0.353	
Peak 1 (burst)	4.505	4.23	0.000	0.17156
Peak 2 ($V0$)	2.618	4.71	0.000	0.21200
Peak 2 ($V50$)	− 2.021	− 4.78	0.000	0.21859
Contrast ($W5$)	3.772	3.60	0.001	0.12427

Acoustic analyses were performed partly to determine if the inconsistent perceptual effects of contrast alterations on /d/ could be based on inconsistent performance of the speech processing software. Direct measures of contrast in normal, enhanced, and reduced stimuli did not indicate a significant failure of the speech processing on /d/ tokens: expected effects of the processing were generally present and significant across all consonants. There was a nonsignificant trend for the differences in contrast between the original and altered /d/ tokens to be smaller than for /b/ and /g/ tokens. However, contrast alterations were in the same direction for all consonants and the differences in magnitude of the physical changes in contrast seem unlikely to account for the large differences in perceptual effects.

The amplitudes of spectral peaks measured at the location of the burst for the original tokens showed a fairly typical pattern of strongest low-frequency amplitude for /b/, strongest midfrequency amplitude for /g/, and strongest high-frequency amplitude for /d/. If these characteristic onset patterns in burst spectra were strong perceptual cues for the listeners, the contrast enhancement procedure used here was particularly biased against /d/. Contrast enhancement tended to increase the amplitudes of peaks in the midfrequencies and reduce the amplitude of the high-frequency peaks thus pushing burst spectra away from those normally associated with /d/. In addition to these amplitude effects, contrast itself was least affected at the frequencies most characteristic of release spectra for /d/. Finally, another methodological factor also may have made identification of /d/ tokens especially difficult for the listeners in this experiment. That was the use of a flat gain characteristic in amplifying stimuli to MCL. For listeners with sloping losses, flat gain entails that higher frequencies are less audible than low frequencies. All these factors are consistent with results reported by Dubno *et al.* (1989) in which stop consonant place confusions were shown to reflect audibility of specific regions of consonant onset spectra.

Taken together, these factors suggest several procedural changes that might result in improved effectiveness of contrast enhancement for /d/ tokens. First, amplification should be tailored to normalize audibility throughout the spectrum (cf., Bustamante and Braida, 1986). Second, contrast alterations should be applied to higher frequencies while continuing to taper alterations at lower frequencies. Third, instead of using the average level of the entire spectrum as a baseline, it may be better to use gross spectrum shape as the baseline against which contrast is computed. That would reduce the tendency for lower amplitude peaks (often at higher frequencies) to actually be further attenuated by contrast enhancement.

Despite the inconsistent effects on /d/, contrast alterations generally had perceptual consequences. To what acoustic factors are the perceptual effects to be attributed? Based on the results of the regression modeling, it was clear that differences in amplitude, and hence audibility, of spectral peaks at consonant release and vowel onset were strongly related to the perceptual effects. In the regression analysis, these amplitude measures accounted for the largest part of the multiple R for the best models. Moreover, the form of the regression models suggested that the important contribution of increased amplitude was mostly apparent during and immediately following the stop consonant release: 50 ms after vowel onset, differences in the amplitude of peak 2 (usually $F2$) were inversely related to differences in percentage correct consonant identification.

But the present data also provide evidence that factors other than simple audibility may have played a role in the perceptual advantage associated with enhanced contrast stimuli. An additional significant proportion of the variance accounted for by the best regression models was attributed to a measure of spectral contrast within the vocalic transition region following consonant release. Thus, after perceptual differences related to amplitude measures were statistically factored from the data, there remained a significant correlation between performance and stimulus contrast. The best models were those that took both amplitude and contrast into account.

This result is consistent with results from a number of recent studies in which audibility alone has not been able to account for the variability in reception scores of hearing-impaired listeners (Pavlovic, 1984; Kamm *et al.*, 1985; Walden *et al.*, 1981; Turner and Robb, 1987). These studies emphasize the fact that hearing loss is associated with both attenuation and distortion of stimulus information. One form of distortion, resulting from a failure of the auditory system to suppress energy surrounding spectral peaks, is poorer frequency resolution. Contrast enhancement, because it effectively carries out a form of suppression at the level of the stimulus, would logically be advantageous for listeners who exhibit poor frequency resolution.

Other studies using similar stimulus manipulations have failed to demonstrate beneficial effects of contrast enhancement for hearing-impaired listeners (Boers, 1980; Summerfield *et al.*, 1985; Dubno and Dorman, 1987; Bustamante and Braida, 1986). Possibly the most important difference between the present study and these others is in the way that contrast was differentially applied to the midfrequency region of the spectrum. As a result, contrast enhancement typically caused a reduction in the amplitude of $F1$ relative to the amplitudes of $F2$ and $F3$. For the present stimuli, had contrast enhancement been applied to the region containing $F1$ as well, an opposite effect would have been obtained with the amplitude of $F1$ increasing to dominate the overall stimulus level. Such an effect could be especially disruptive since $F1$ has been implicated in the masking of higher formants for hearing impaired listeners (Danaher and Pickett, 1975). Thus stimulus manipulations that result in increasing the prominence of $F1$ may be perceptually disruptive for many hearing-impaired listeners. These two factors, the relative attenuation of higher formants and increased upward spread of masking from $F1$, could account for differences between the present and previous studies.

In summary, the present results suggest that for hearing-impaired listeners some perceptual advantage might be gained from enhancement of spectral contrast. However, this conclusion is quite tentative. Further tests of the effectiveness of this processing will require successful application to a much wider range of stimuli. The particular processing scheme used here is computationally expensive, especially when compared to simpler frequency-dependent amplification schemes. Its practicality as a speech enhancement for hearing-impaired listeners, say in a digital signal processing hearing aid, rests largely on continuing to demonstrate perceptual advantages beyond those that can be achieved via amplification alone.

ACKNOWLEDGMENTS

This work was conducted while the author was a Research Scientist at Gallaudet University in the Center for Auditory and Speech Sciences. Work was supported by NIDCD Grant Nos. NS23249 (to the University of Maryland) and NS05464 (to Gallaudet University). The author would like to thank James M. Pickett, Sally G. Revoile, and reviewers Diane Bustamante and Chris Turner for their many helpful comments. The assistance of Nicola Pate and Doris Bitler in data collection and analysis is gratefully acknowledged.

[1] In addition to the reported analyses, similar analyses were carried out using an arcsine transformation to the raw proportion correct data. The *F* ratios tended to be slightly larger for the arcsine transformed data, however, differences resulting from data transformation did not alter interpretation of the results and consequently the presentation of the results is based on analyses of the raw data.

Blumstein, S. E., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017.

Boers, P. M. (**1980**). "Formant enhancement of speech for listeners with sensorineural hearing loss," in *IPO Annual Progress Report*, No. 15 (Institut voor Perceptie Onderzoek, The Netherlands), pp. 21–28.

Bunnell, H. T., and Martin, J. G. (**1988**). "Acoustic correlates of intervocalic stop confusions," J. Acoust. Soc. Am. Suppl. 1 **84**, S158.

Bustamante, D. K., and Braida, L. D. (**1986**). "Wideband compression and spectral sharpening for hearing-impaired listeners," J. Acoust. Soc. Am. Suppl. 1 **80**, S12–S13.

Bustamante, D. K., and Braida, L. D. (**1987**). "Principal-component amplitude compression for the hearing impaired," J. Acoust. Soc. Am. **82**, 1227–1242.

Danaher, E. M., and Pickett, J. M. (**1975**). "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss," J. Speech Hear. Res. **18**, 261–271.

Daniel, C., and Wood, F. S. (**1971**). *Fitting Equations to Data* (Wiley, New York), p. 86.

Dubno, J. R., and Dorman, M. F. (**1987**). "Effects of spectral flattening on vowel identification," J. Acoust. Soc. Am. **82**, 1503–1511.

Dubno, J. R., Dirks, D. D., and Ellison, D. E. (**1989**). "Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. I: The contribution of selected frequency regions," J. Acoust. Soc. Am. **85**, 347–354.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). "Speech recognition and the articulation index for normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 281–288.

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech-reception threshold for sentences," Audiology **18**, 43–52.

Summerfield, Q., Foster, J., and Tyler, R. (**1985**). Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants," Speech Commun. **4**, 213–229.

Turner, C. W., and Robb, M. P. (**1987**). "Audibility and recognition of stop consonants in normal and hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1566–1573.

Tyler, R. S., Fernandes, M., and Wood, E. J. (**1980**). "Masking, temporal integration and speech intelligibility in individuals with noise-induced hearing loss," in *Disorders of Auditory Function*, edited by I. Taylor and A. Markides (Academic, London), Vol. III.

Walden, B. E., Schwartz, D. M. Montgomery, A. A., and Prosek, R. A. (**1981**). "A comparison of the effects of hearing impairment and acoustic filtering on consonant recognition," J. Speech Hear. Res. **24**, 32–43.

Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am. **82**, 1548–1559.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.