Research Article

Vowel Identification by Amplitude and Phase Contrast

MICHELLE R. MOLIS^{1,2}, ANNA DIEDESCH^{1,3}, FREDERICK GALLUN^{1,2}, AND MARJORIE R. LEEK^{1,2}

¹VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), Portland VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, OR 97207, USA

²Oregon Hearing Research Center, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA

³Present Address: Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA 98105, USA

Received: 19 December 2011; Accepted: 11 September 2012; Online publication: 25 September 2012

ABSTRACT

Vowel identification is largely dependent on listeners' access to the frequency of two or three peaks in the amplitude spectrum. Earlier work has demonstrated that, whereas normal-hearing listeners can identify harmonic complexes with vowel-like spectral shapes even with very little amplitude contrast between "formant" components and remaining harmonic components, listeners with hearing loss require greater amplitude differences. This is likely the result of the poor frequency resolution that often accompanies hearing loss. Here, we describe an additional acoustic dimension for emphasizing formant versus non-formant harmonics that may supplement amplitude contrast information. The purpose of this study was to determine whether listeners were able to identify "vowel-like" sounds using temporal (component phase) contrast, which may be less affected by cochlear loss than spectral cues, and whether overall identification improves when congruent temporal and spectral information are provided together. Five normal-hearing and five hearing-impaired listeners identified three vowels over many presentations. Harmonics representing

formant peaks were varied in amplitude, phase, or a combination of both. In addition to requiring less amplitude contrast, normal-hearing listeners could accurately identify the sounds with less phase contrast than required by people with hearing loss. However, both normal-hearing and hearing-impaired groups demonstrated the ability to identify vowel-like sounds based solely on component phase shifts, with no amplitude contrast information, and they also showed improved performance when congruent phase and amplitude cues were combined. For nearly all listeners, the combination of spectral and temporal information improved identification in comparison to either dimension alone.

Keywords: spectral contrast, formants, hearing impairment, frequency selectivity, phase, vowel identification

Portions of this research were presented at the 2008 meeting of the American Auditory Society, Scottsdale, AZ, and at the 2008 International Hearing Aid Conference, Lake Tahoe, CA.

Correspondence to: Michelle R. Molis • VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), Portland VA Medical Center • 3710 SW US Veterans Hospital Road, Portland, OR 97207, USA. Telephone: +1-503-2208262; fax: +1-503-7211402; email: michelle.molis@va.gov

INTRODUCTION

Increasing the spectral contrast (amplitude difference) between formant frequency peaks and valleys in speech spectra has been studied as a way to improve speech intelligibility in hearing-impaired listeners, but with only small success (see, e.g., Simpson et al. 1990). In fact, too much spectral contrast enhancement may begin to distort the intended signal (Baer et al. 1993). DiGiovanni et al. (2005) wondered whether spectral enhancement was a viable signal processing approach to improving speech recognition by people with hearing loss. Using

simplified stimuli and psychophysical measurements, the authors demonstrated that many hearing-impaired listeners can improve their auditory discrimination and detection with increased differences in spectral peaks and valleys, leading them to conclude that spectral enhancement may provide some benefit for improved speech recognition, especially in noise backgrounds.

The ubiquitous difficulty of enhancing spectral contrast in stimuli that are to be processed through auditory filter banks with abnormally broad spectral channels was noted by Giguère and Smoorenburg (1998), who pointed out that increasing peaks and decreasing valleys in a speech spectrum was unlikely to provide benefit to hearing-impaired persons because of their abnormal frequency selectivity. They suggested, however, that the exploitation of some temporal aspects of auditory processing might be able to enhance spectral peaks by increasing neural synchrony, as was shown physiologically by Miller et al. (1999). Perhaps the addition of temporal or phase contrast to spectrally coded peaks and valleys in a speech spectrum might add to the enhancement effect independent of amplitude contrasts. Typically, phase is disregarded in speech signals because it is lost due to acoustical reflections in the environment, but it is possible to preserve and to detect phase changes under headphones, or even through hearing aids. Coding speech information through the manipulation of phase may enhance the acoustical cues used to identify some speech sounds when spectral contrasts are degraded by hearing loss. The combination of both temporal and spectral information might improve the identification of some speech sounds.

In earlier work focused on the loss of spectral contrast in hearing-impaired listeners, Leek et al. (1987), Leek and Summers (1996), Alcántara and Moore (1995), Alcántara et al. (1996), and Dreisbach et al. (2005) created synthesized “vowels” by varying the component amplitudes of harmonic complexes to highlight “formant” frequency regions over background harmonics. For instance, Leek et al. (1987) used a harmonic complex with a fundamental frequency of 100 Hz and 36 components, all in cosine phase, to create four such vowels. Normal-hearing listeners were able to correctly identify the vowels with only 2 dB of increased formant amplitude (spectral contrast), while hearing-impaired listeners required at least 6 dB to perform at the same level. Traunmüller (1987) constructed vowel-like sounds using a harmonic complex, but varying only the component phase of the formant frequency regions by 180° relative to the phases of the remaining components. With no amplitude contrast provided, normal-hearing listeners were able to accurately identify nine “phase vowels” with better than 50 % accuracy (chance being 11 %) when the harmonic complexes had a low

fundamental frequency (71 Hz). Traunmüller’s work showed that component phases do have a measurable impact on stimulus identification which ordinarily is based primarily on amplitude contrasts.

The purpose of this study was to code vowel-like stimuli with phase information to determine whether the component phase could be used by normal-hearing (NH) and hearing-impaired (HI) listeners for vowel recognition. The identification of vowel-like harmonic complexes was measured as a function of formant amplitude and formant phase. The stimuli in each of two conditions were constrained to have only relative amplitude information at the formant frequencies (i.e., equal phase across components) or relative phase information at formant frequencies (i.e., equal amplitude across components). In a third condition, the component phase and amplitude enhancements were combined to test whether adding phase contrast would improve identification based on amplitude contrast alone.

METHODS

Listeners

Five NH listeners and five HI listeners served as the subjects in this study. NH listeners ranged in age between 27 and 60 years ($M=48.0$ years) and had hearing thresholds of 25 dB HL or better at audiometric frequencies from 250 to 4,000 Hz. HI listeners ranged in age between 61 and 74 years ($M=65.2$ years) and had relatively flat sensorineural hearing losses of at least 30 dB HL in the low frequencies (<1,000 Hz) and 40–60 dB HL in the mid and high frequencies (1,000–6,000 Hz). Comparisons of air and bone conduction audiometric thresholds, along with emittance results confirmed the sensorineural nature of the hearing losses. Figure 1 shows individual air conduction thresholds for the HI listeners and the average thresholds of the NH group, along with standard errors of the mean. Testing was accomplished in three to four 2-h sessions. All subjects provided written informed consent prior to participation and were paid per session.

Stimuli

The stimuli for this study were harmonic complexes comprising the first 36 harmonics of a 100-Hz fundamental frequency. Simulated vowels were created by adjusting the amplitudes and/or phases of three pairs of components within the harmonic complex at the locations of three “formants” chosen to approximate the average formant frequencies of male speakers for the vowels /i/, /α/, and /æ/ (Peterson and Barney 1952). Table 1 lists the frequencies of the altered harmonic pairs for each of the three vowel

FIG. 1. Air conduction thresholds (in decibels HL re: ANSI 2004) for the listeners in this study. The average NH thresholds, with standard errors of the mean, are indicated by *closed blue symbols*; individual thresholds for the HI listeners are indicated by *open red symbols*.

categories. All of the altered components were changed by the same amount within a stimulus.

Three different types of vowel-like stimuli were created: those for which only the amplitudes of the formant harmonics were altered (amplitude contrast), those for which only the phases of the formant harmonics were altered (phase contrast), and those for which both amplitudes and phases were altered (combined contrast). For the amplitude contrast condition, the level of the appropriate three pairs of harmonic components was increased from 1 to 10 dB above the level of the 30 background harmonics following procedures similar to those of Leek et al. (1987). All harmonics had a cosine (0°) starting phase. In the phase contrast condition, all harmonics were equal in amplitude. All background harmonics had a cosine starting phase and the formant harmonics were shifted in phase relative to cosine by 4 – 180° . The degree of phase shift was the same for all formant harmonics within a given stimulus. In the combined contrast condition, both the amplitude and phase of the formant harmonics were altered; the background components maintained equal amplitude and cosine phase. Further details on the amplitude and phase increments used in the experimental tasks are provided in the following sections.

Stimuli were generated digitally at a rate of 24,414 samples per second. Each stimulus was 400 ms in

duration, with 50-ms cosine-squared onset and offset ramps. The root mean square (rms) values of these waveforms change with alterations in the component amplitude, although not with phase changes alone. As part of the stimulus generation, all waveforms were normalized to the same rms value. Figure 2 schematically shows examples of stimulus amplitude and phase spectra (left panels) and 40-ms samples of the corresponding stimulus waveforms (right panels) for the vowel /i/ for the amplitude-only (top), phase-only (middle), and combined (bottom) conditions. The example shows the spectra and waveforms for the vowel /i/, with either 6 dB of amplitude increment, shown by the bottom amplitude spectrum referenced to the left-hand axes, and/or a 180° phase shift of the formant harmonics, shown by the upper phase spectrum in each panel referred to the right-hand axes. The waveforms shown in the right column of panels correspond to each amplitude and phase combination.

Procedures

Listeners were tested individually in a sound-treated booth. The stimuli were presented monaurally via Sennheiser HD 580 precision headphones at an overall level of 85 dB SPL with up to a ± 3 -dB random level rove on each presentation to defeat possible loudness differences (Gockel et al. 2003). NH listeners were tested in their better ear and HI listeners were tested in the ear that best matched the flat audiometric criteria.

Data were collected in three separate identification experiments: vowel identification based on adjustments in the component phase alone, vowel identification based on adjustments in the component amplitude alone, and vowel identification based on adjustments in both phase and amplitude components changing together. Testing began with the phase contrast task, next moving to the amplitude contrast task, and finishing with the combined task. However, if a listener was unable to reach criterion performance on the phase contrast training task (described below), the amplitude contrast task was completed first and then the phase contrast task was revisited. This occurred for two NH listeners and two HI listeners.

A single-interval three-alternative forced choice identification procedure was used, with correct answer feedback. Stimuli within the blocks were presented in random order. On each trial, listeners were asked to identify a stimulus by pressing one of three response buttons displayed on a touch screen monitor. The buttons were labeled “ee,” “ah,” and “ae” for the response categories /i/, /α/, and /æ/, respectively. Directly below the touch screen buttons, the key words “heat,” “hot,” and “hat” were also displayed.

Prior to testing for each condition, a training task was carried out which consisted of a single change in

TABLE 1

Stimulus “formant” harmonic frequencies for each of three simulated vowels based on Peterson and Barney (1952)

	/i/	/α/	/æ/
F1	200, 300	700, 800	600, 700
F2	2,200, 2,300	1,000, 1,100	1,700, 1,800
F3	3,000, 3,100	2,400, 2,500	2,400, 2,500

FIG. 2. Schematics of amplitude and phase spectra (*left*) and 40-ms samples of the corresponding stimulus waveforms (*right*) for the /i/ response category. The *top row* displays an amplitude-only incremented stimulus, the *middle row* displays a phase-only incremented

stimulus, and the *bottom row* displays a stimulus for which both amplitude and phase have been incremented. *Left-hand panels* show amplitude spectra referred to the left ordinate and phase spectra referred to the right ordinate.

only one stimulus property of the formant harmonics (either a 180 ° phase change or a 5- or 10-dB amplitude increment over the background harmonics, depending on the condition to be tested). During the training task, the listeners were asked to identify the stimuli, with feedback provided after each response. This was repeated for up to five blocks of ten replications per block (30 trials per block). However,

if after three blocks the listener demonstrated that he or she could correctly identify the simulated vowels (>90 % correct in all of the first three blocks), testing would begin. For the phase contrast task, the listener had to achieve a score of at least 80 % correct in one of the five blocks during the training to continue with the testing. Otherwise, that listener would move on to the amplitude contrast task and be given a second

chance with the component phase task upon completion of the component amplitude testing. One hearing-impaired listener who failed to meet the 80 % correct phase criterion on initial testing was still unable to meet that criterion after completing the amplitude task. That listener was excluded from further testing and was replaced by another hearing-impaired subject. There were no obvious auditory or demographic characteristics that would set this listener apart from the others enrolled in the study. All other subjects were able to correctly identify vowels based on the 180 ° phase shift alone.

For the initial block of the phase contrast task, the phases of the formant components were shifted from the background components by equally spaced increments ranging from 36 to 180 °. In each block, five phase increment values were selected within that range and ten repetitions were presented for each of the three “vowels” (150 trials per block). If the listener obtained 90 % correct at each of these five phase increments, the formant phases were reduced after three test blocks. This occurred for two NH listeners and no HI listener. For the other listeners, data were collected for five blocks. After the data were obtained for this initial range of phase contrasts, the component phase increments were then further reduced in small steps, with ten replications at each step and vowel sound until chance levels of performance were approached. Chance, or guessing percentage, was defined as 33 % correct for the three-alternative identification task. To verify chance-level performance in this probabilistic task, a probability distribution for the range of scores for chance performance was computed by running a simulation of random responding for 10,000 sessions of 20-trial blocks. The results of the simulation indicated that a listener performing at chance levels would achieve an overall mean percent correct score of 33 % with a 95 % confidence interval of ± 5 %. Therefore, any block for which percent correct was < 38 % was taken to be chance performance. Over the total range of levels tested, the phase increment that produced approximately 67 % correct identification performance (midway between chance and perfect identification) was selected as the phase contrast threshold. Phase values near these threshold increments were subsequently used in the combined phase and amplitude experiment.

The amplitude contrast identification (i.e., spectral contrast) was presented in a manner similar to the component phase experiment, but the amplitudes of harmonic pairs were varied rather than the phases. Based on the findings of Leek et al. (1987), it was anticipated that NH listeners would need approximately 2 dB of amplitude contrast in order to achieve about 67 % correct and that HI listeners would require at least 6 dB of contrast. With this in mind, NH listeners were tested in 1-dB steps ranging from 1 to 5 dB of

component amplitude and HI listeners were tested first in 2-dB steps ranging from 2 to 10 dB, and, subsequently, also in 1-dB steps ranging from 1 to 5 dB.

Finally, in the combined task, two component amplitude increments (always 1 and 2 dB) were combined with three component phase variants that were chosen individually for each listener based on his or her previous performance on the phase contrast task. All listeners were tested at 1- and 2-dB amplitude contrast levels and at customized phase increments that were near threshold performance and slightly above. The phase increments used across listeners varied from 4 to 45 °.

Data for each of the three tasks were percent correct vowel identifications. Error bars on the graphical displays of the data indicate standard errors of the mean. Statistical analyses were implemented using the software package SYSTAT®. Details of these analyses are described in “Results.”

RESULTS

In each experiment (amplitude, phase, and combined tasks), formant peak contrast levels along the given dimension(s) were varied and percent correct identifications were determined for each level.

Amplitude contrast task

The main purpose of testing the amplitude-alone spectral contrast condition was to determine the base amplitude increment for the subsequent combined amplitude and phase contrast condition. It was expected, based on a considerable amount of earlier work, that there would be small differences in the amount of amplitude contrast required for vowel identification by NH and HI listeners.

Figure 3 displays the mean performance across the two listener groups for increments in the formant component amplitudes. Error bars represent standard errors of the mean. All listeners in both listener groups heard the same range of amplitude increments. Note that performance improved to near perfect identification for all listeners over the range of the spectral contrast values used here. As has been reported previously (e.g., Leek et al. 1987), HI listeners required slightly more spectral contrast than NH listeners in order to perform at the same level.

On average, better-than-chance performance (> 38 % correct) was observed for all contrasts for all listeners. Contrasts of about 1.5 dB were required by NH listeners and just over 2 dB for HI listeners in order to achieve an arbitrary threshold level of about 67 % correct. An analysis of variance for this condition, with a between-subjects factor of group

FIG. 3. Mean percent correct identification and standard error of the mean for NH and HI listener groups (filled blue and open red symbols, respectively) for increments in component amplitude.

membership and a repeated measures factor of amplitude contrast, indicated that there was a significant difference due to contrast level [$F(4,32)=189.23$, $P<0.0001$] and a significant contrast-by-hearing status interaction [$F(4,32)=72.08$, $P=0.01$]. There was not a significant difference in overall performance between the NH and HI listeners ($P=0.27$). The significant interaction reflects the approach to an asymptotic value of performance for a lower spectral contrast value for the NH listeners relative to the HI listeners.

A replication of this condition was carried out after all other conditions were completed as a control condition in order to verify that any improvements due to the combined phase and amplitude enhancements were not simply due to additional experience with these tasks. Recall that the combined phase and amplitude condition was always presented after the individual phase and amplitude conditions were completed. Any benefits observed from adding coherent phase information in the combined task might be attributed to improvements in performance on the amplitude task alone after practice; that is, if listeners improved due to practice with these conditions, the contribution of the added phase information might be minimized. For this reason, the amplitude-alone condition was presented again after all the other data had been collected. The results of the replicated amplitude-only condition were nearly identical to the initial data, confirming that any improved performance observed in the combined phase and amplitude condition would indeed represent a contribution from phase information combined with the amplitude contrast information.

Phase contrast task

In the phase contrast condition, all component amplitudes were held constant, and the formant components were encoded by shifting the phases relative to the

phases of the background components. Recall that the procedures used did not require all listeners to be tested at the same phase values, but instead the phases were selected for each subject, systematically decreasing until the discrimination was no longer better than chance performance (see “Methods”). Figure 4 shows identification for each NH and HI listener for a range of phase values. NH listeners are shown with solid symbols and HI listeners with open symbols, and each subject’s data are coded with a different shape. The two lines on Figure 4 show the best-fitting logistic dose–response curve to the data from each group of listeners (NH, solid line; HI, dashed line).

Normal-hearing listeners were generally better overall at using the phase increments for vowel identification, but both groups showed strong improvements as the phase shifts moved from about 10° to near 100° . Both groups showed a similar amount of improvement across the entire range of phase increments tested and approached an asymptotic performance level (a mean of about 90 % correct for NH listeners and about 80 % for HI listeners).

Figure 5 displays these logistic fits to data from the individual subjects. The logistic fits were used to extract three values for statistical comparison: phase threshold, asymptote, and slope. The phase threshold, defined as midway between chance (33 %) and perfect performance (100 %), or 67 % correct, did not differ statistically between the two groups [$t(8)=$

FIG. 4. Observed (symbols) and fitted (lines) vowel identification scores for the phase-only condition (amplitude increments=0) for NH listeners (filled blue symbols) and HI listeners (open red symbols). The solid blue line indicates the average fitted function for the NH listeners; the dashed red line indicates the average fitted function for the HI listeners.

FIG. 5. Fitted vowel recognition scores as a function of phase increments alone for individual NH and HI listeners (*left and right panels, respectively*).

-1.81, $P=0.32$], nor did the estimated asymptotic performance level [$t(8)=1.12$, $P=0.89$]. The slopes of the fitted functions were significantly steeper for the HI group than for the NH group [$t(8)=5.05$, $P=0.003$] (all P values were Bonferroni-corrected for multiple tests). Although identification performance was lower for the HI listeners at the small phase increments, their performance improved at a faster rate than for the normal-hearing group such that, as a group, the asymptotic levels were similar. These data indicate that both NH and HI listeners can use phase information alone (i.e., without any spectral amplitude contrast) to perform these vowel identifications.

Phase and amplitude contrast combined task

In addition to the phase-only condition, with no formant information provided in the amplitude spectrum of the stimuli, identification was also measured for stimuli that had small amplitude increments on the formant harmonics that were below- or near-amplitude contrast thresholds, combined with phase shifts that were in the range of the phase-only thresholds for each individual. These phase selections fell from below the phase-only threshold to just above the threshold contrasts for individual listeners, chosen by visual inspection of the individual listeners' phase contrast psychometric functions (see Fig. 5). These phase contrasts were combined with amplitude contrasts of 0, 1, and 2 dB. This experiment was included to determine whether identification performance could be improved by combining sub-threshold amplitude and phase increment values, neither of which alone would support vowel identification.

Although all the phase values used in the combined experiment were within a range between 4 and 45°, the same values were not used for every subject; therefore, an analysis of variance (ANOVA) with the same phase increment for each subject was precluded.

In order to allow an appropriate factorial ANOVA, the data for each subject were fitted separately for each of the three amplitude conditions (0-, 1-, and 2-dB amplitude increments). The logistic dose-response used for the phase-alone experiment was again applied to the individual data here. In order to fit the ogive to the individual data, a "dummy" point was included at 100 % correct at 100° of phase, but the actual data points between 0 and 40° were weighted more than the dummy point. This process produced fitted functions whose R^2 goodness-of-fit values were at least 0.99 for 23 of the 30 fits, with the worst fit having a goodness of fit of 0.94. These fitted functions were entered to estimate performance at phases of 0, 10, 20, and 30° for each of the three amplitude values, and these fitted values were analyzed with a repeated-measures ANOVA, with a between-subjects factor of hearing status and repeated measures on amplitude and phase increments.

Figure 6 shows the estimated percent correct at each of the selected phase values for the three amplitude values, with separate panels for the NH and HI data. Across amplitudes, the mean performance thresholds in the HI group are generally lower than for the NH group. The main effect of group membership on the performance levels was significant [$F(1,8)=5.75$, $P=0.04$]. There was also a significant main effect of amplitude value [$F(2,16)=89.68$, $P<0.001$] and of phase [$F(3,24)=96.74$, $P<0.001$]. Significant interactions were also identified between phase value and hearing group membership [$F(3,24)=6.53$, $P=0.002$] and amplitude by phase [$F(6,48)=7.29$, $P<0.001$]. Amplitude by hearing status was not a significant interaction ($P=0.86$), but the three-way interaction between group, amplitude, and phase was significant [$F(6,48)=6.44$, $P<0.001$]. These statistical findings indicate that there were differences in identification performance for the phase and amplitude combinations over the measurements here.

FIG. 6. Estimated percent correct and standard errors for combinations of three amplitude contrasts (0, 1, and 2 dB) and four phase contrasts (0, 10, 20, and 30 °) for NH and HI listeners (*left and right panels*, respectively). Data were extracted from individual listeners' fitted functions to the percent correct identification at each of the three amplitude contrasts.

However, some of those differences involving group membership may be related to ceiling effects in the NH data.

The three-way interaction suggests that improvements resulting from the combined effects of amplitude and phase contrast were not the same for both groups. In fact, inspection of Figure 6 suggests that, for normal-hearing listeners (*left panel*), once the amplitude increment was at least 2 dB (top data series, triangles), any contribution of phase increment was essentially swamped by the amplitude cue, so additional phase information resulted in smaller additional performance improvements. The HI listeners (*right panel*) did not demonstrate ceiling effects, and so their performance continued to show improvements. Figure 7 shows the average maximum improvements for each group and amplitude condition. These values show the percentage point differences for each group and condition when the maximum phase (30 °, taken from the fitted curves used for Fig. 6) was added to each amplitude condition. For the NH group,

FIG. 7. Mean maximum improvement in estimated percent correct achieved with an increase of phase contrast from 0 to 30 ° for the two listener groups for each of three amplitude contrasts (0, 1, and 2 dB).

adding phase information to a flat-amplitude spectrum (the 0-dB condition) improved performance by 50 percentage points. Improvements to the 2-dB condition were, however, only about 14 percentage points. The HI group's performance improved by about 20 percentage points for each amplitude condition. This figure shows that, even for sub-threshold spectral peak-to-valley contrasts, as might occur in hearing-impaired listeners due to a smearing of their internal spectra, the addition of a limited amount of formant phase contrast could potentially enhance identification performance by 20 percentage points.

DISCUSSION

The results of this study demonstrate that the location of "formants" in stylized vowel-like stimuli can be coded by either incrementing the amplitudes of the relevant harmonics or by shifting the phases relative to the background harmonics for both NH and HI listeners. In addition, the combination of amplitude and phase information improved identification performance for both listener groups over either dimension alone. Nearly all listeners received benefit from the addition of coherent phase information in the locations of near- or sub-threshold amplitude contrast.

NH listeners required less amplitude (spectral contrast) than HI listeners in order to identify these vowel-like sounds. This is consistent with previous findings (e.g., Leek et al. 1987), but in the current study, the differences between groups were minimal—between 1 and 2 dB of contrast was required for NH listeners and between 2 and 3 dB for comparable performance by HI listeners. This small amount of difference is probably related to the limited stimulus set, with only three different response categories instead of four. NH listeners, in general, could take slightly better advantage of formants cued by compo-

nent phase alone than HI listeners (Fig. 4), especially for small phase increments; however, both groups of listeners were highly variable, and there was a considerable overlap between the two groups in the range of asymptotic performance achieved (Fig. 5).

Most listeners showed improvements in the combined phase and amplitude condition, but the benefit varied widely across listeners. The size of the improvement realized by HI listeners for added phase information was roughly consistent across differing levels of amplitude contrast. The retesting of the amplitude-alone condition after the combined conditions confirmed that the improved performance observed from adding phase contrasts to amplitude contrast was not simply due to more practice with the amplitude condition, but represented a true identification enhancement contributed by the addition of phase information.

Although the responses in this study were fairly simple (one of three vowel sounds), it is possible that any observed differences in performance between the NH and HI listeners could potentially be attributable to the differences in age between the listener groups as well as the audibility of the stimuli. The two groups of listeners differed in age; however, there was not a relationship within the hearing-impaired group between age and any of the performance measures in this study. Additionally, it should be noted that the differences observed between the two groups could not be attributed to an inability of the hearing-impaired group to perform the task, given enough amplitude and/or phase increments. In fact, statistically, there were few differences between groups on measures of overall performance, with interactions between increment and group membership on both the amplitude and phase conditions relating to how rapidly subjects rose from near chance to asymptotic level performance. There was a group difference in performance for the combined condition, but this was likely related primarily to the approach to ceiling performance by the NH listeners, but not the HI listeners. If age were a significant contributor to these results, the overall performance levels would likely have been considerably lower for the HI listeners.

There may have been a difference in audibility of the stimuli for the two groups of listeners. The presentation levels were high (around 85 dB SPL), but it is possible that some subjects in the hearing-impaired group may not have been able to hear some of the higher frequency formants. To investigate this possibility, an excitation pattern analysis was undertaken following procedures outlined by Moore and Glasberg (2004) and Moore (2007). Because audibility would be affected by the amplitude spectrum, not the phase spectrum, the analysis was done only on the flat-amplitude spectrum. Any formant increments would

be added to this flat spectrum, so the strictest test for stimulus audibility would be spectra with no amplitude increments. The audiometric thresholds for each hearing-impaired subject were entered into the calculation developed by Moore and Glasberg relating the broadening of auditory filters to impaired thresholds (see Eq. 13, p. 78, of Moore and Glasberg). Starting with “normal” bandwidth auditory filters, described by Moore and Glasberg, the auditory filter broadening was applied at each audiometric frequency for each subject’s hearing loss to construct an auditory filter bank specific to that subject. This filter bank was then used to estimate the minimum amount of excitation required at each audiometric frequency in order for a stimulus to be audible. The dB value representing the difference between that minimum threshold excitation and the excitation produced by the flat-amplitude harmonic spectrum was taken as a measure of the audibility of the stimulus at each formant frequency region for all three vowel stimuli for each subject. The results of this excitation pattern analysis were that for all hearing-impaired subjects, the frequency regions around the first two formants were presented at between 15- and 30-dB sensation level (SL). The third formant frequency region was about 15-dB SL for all subjects, except for HI10 for which it was closer to 5-dB SL. This did not seem to negatively impact this subject as Figure 5 shows this subject to be one of the most successful performers in the phase-alone task.

A further analysis regarding the potential for some of the frequency regions to be less than audible was undertaken by viewing the confusions among the three vowel responses. If some frequency regions were inaudible, it would be expected that greater confusions would exist for some pairs of stimuli relative to other kinds of errors. This was not the case. Across all conditions and for both listener groups, no consistent error patterns or confusions between vowels were observed. Errors were approximately equally distributed within response categories and did not appear to be dependent on listener characteristics or group membership.

Amplitude and phase contrasts simulated in a model of auditory processing

To gain insight into how component phase and amplitude contrasts may be encoded in both normal and impaired auditory systems, the auditory image model (AIM), developed by Patterson and his colleagues (Patterson et al. 1992, 1995; Patterson 2000), was applied to the stimuli used in this study. The AIM is a time-domain model of auditory processing that preserves the details of temporal characteristics of a signal in its output. Successive stages in the auditory

processing pathway are simulated in separate processing modules. The modules of the AIM employed here simulated the effects of pre-cochlear processing, basilar membrane motion, and neural activity patterns. The reader is referred to Bleeck et al. (2004), as well as Patterson (1994), for further details concerning the AIM and its application.

Stimuli used in this modeling were all three vowel categories, with all amplitude conditions (1–5 dB of amplitude contrast), several phase contrast conditions ranging from 5 to 180°, plus the 0 phase and 0 amplitude (flat spectrum) conditions as references. Combined amplitude and phase conditions were simulated using 1- and 2-dB amplitude contrasts and 30° phase contrast. Modeling simulated both normal and impaired auditory processing.

In the first module (pre-cochlear processing), an equal-loudness contour compensation was applied to the signals using data from Glasberg and Moore (2002). The outputs of that module were passed to the basilar membrane module, which simulated a bank of 75 linear “gammatone” filters (Patterson et al. 1995). The standard filter widths used to model the response of an unimpaired basilar membrane are based on Glasberg and Moore (1990). For the impaired version of the model, broadening of the auditory filters for the simulation of impaired frequency selectivity was achieved by linearly increasing the filter bandwidths as a function of processing channel by altering the appropriate parameter in the gammatone filter calculation within the AIM software. As described earlier for the audibility analysis, Moore and Glasberg (2004) estimated that, in general, auditory filters broaden according to increasing absolute thresholds. The impaired filter bank simulated here for the AIM represents an approximation to the broadening filter bandwidths typically observed as the absolute thresholds of most hearing-impaired listeners increase with frequency. For the lowest frequency channel (centered at 100 Hz), the width of the filter was equal to that of a standard unimpaired filter, but by the highest frequency channel (6,000 Hz), the filter width was about 2.35 times greater than the standard value. Lastly, the neural activity pattern resulting from cochlear processing was applied to the output of the basilar membrane module. This processing was carried out separately for each of the 75 filter channel outputs and included half-wave rectification, square root compression, and low-pass filtering at 1,200 Hz to simulate the gradual loss of phase locking at higher frequencies. Note that this low-pass filtering resulted in smoothing of the within-channel waveforms for each center frequency channel as the temporal fine structure above 1,200 Hz was progressively removed from each channel.

The output of this sequence of modules consisted of 75 simulated channels of (internal) temporal representations, preserving the differences due to component phases, not just amplitude differences in the stimuli. The average values across time for each frequency channel may be combined into a frequency profile, similar to an excitation pattern representation, which displays estimated excitation at each center frequency. These excitation patterns were compared across all conditions.

The changes in model output associated with incrementing the amplitudes of selected components that code the vowel / α / in an otherwise equal-amplitude, equal-phase harmonic complex are shown in Figure 8 (the other two vowel categories resulted in similar excitation patterns, with peaks at the incremented harmonics specific to their vowel identity; only the / α / category will be shown here for economy of space). The excitation pattern in response to the flat-amplitude stimulus is shown as a thin dashed line. The bar graph along the bottom of the figure indicates the frequencies at which components were incremented in amplitude. Not unexpectedly, the outputs of the models increased as the amplitudes were incremented. The output of the NH model (left panel) demonstrates narrower peaks and lower peak amplitudes than does the output of the HI model (right panel). This reflects the broadening of the filters for the HI model in that each filter sums the energy across more harmonics than observed for a normal filter bank.

The model outputs associated with incrementing the phases of selected components to code the vowel / α / in an otherwise equal-amplitude, equal-phase harmonic complex are shown in Figure 9. Again, the reference condition of no phase or amplitude increments is shown by the thin dashed line. The line graph along the bottom of the figure indicates the frequencies at which components were incremented in phase. Five increments between 13 and 180° are shown. As in the amplitude increment models, the NH version (left panel) shows lower overall outputs than the modeled HI representation, as well as somewhat narrower peaks in response to formant harmonics. As with the amplitude increments, the differences in model output between hearing status groups increase as the frequency increases due to the impaired frequency selectivity for the HI model. Because the AIM preserves the time domain characteristics of the stimuli, differential increases in the output of the model were observed mainly in frequency channels where harmonic components were at least partially unresolved. Generally, for both NH and HI models, growth of the output in channels with more unresolved components increases faster than output in channels with fewer unresolved compo-

FIG. 8. Changes in AIM output produced by several increments in “formant” amplitudes for a model with standard (*left panel*) and impaired (*right panel*) auditory filter bandwidths. Lower bar graphs show a schematic example of the amplitude spectrum that was modeled.

nents. Nevertheless, it is clear from Figures 8 and 9 that both amplitude and phase contrasts have a similar strong effect on the internal representations of these vowels, suggesting that both experimental manipulations may serve to produce peaks at appropriate frequency regions for vowel identification. Simulations of the combined conditions were also generated using the AIM. As expected, combining increments in amplitude and phase increases the excitation peaks more than for either cue alone for those frequency channels containing unresolved harmonics. This is reflected in the improvements observed behaviorally when sub-threshold amplitude and phase cues are included in a vowel stimulus. Figure 10 (left panel) shows behavioral data, both singly and in combination, for amplitude and phase increments that separately are either sub-threshold (1 dB and 10°) or supra-threshold (2 dB and 30°) for both listener groups. In each case, and for both NH and HI listeners, identification performance for the combined stimulus exceeds that of either amplitude or phase increments alone.

Listeners’ identification accuracy was compared to the summed AIM output in the region of the second formant for the same stimulus conditions. For each vowel category, four output channels spanning the frequencies of the two F2 harmonics were selected and their outputs were summed. Likewise, four channel outputs from adjacent regions of spectral valleys in the excitation patterns were summed. Differences were then calculated for the F2 region of each vowel relative to that for the “valley” channels to produce a rough estimate of a peak-to-valley contrast in the internal spectral representation. These estimates are shown in Figure 10 (right panel). As for the behavioral data, the combined stimuli produced consistently greater values. Unlike the behavioral data, however, the model output was consistently greater for the HI than for the NH, once again reflecting the broader auditory filters of the former group. This is a reminder that the magnitude of the formant peaks as well as increases in the peak-to-valley differences may not be sufficient for vowel identification if the peaks are very broad, such as they are in this simulation of an impaired excitation pattern.

FIG. 9. Changes in AIM output produced by increments in “formant” phase for a model with standard (*left panel*) and impaired (*right panel*) auditory filter bandwidths. Lower line graphs show a schematic example of the phase spectrum that was modeled.

FIG. 10. *Left panel*, Mean percent correct identification for combined amplitude and phase increments that separately are either sub-threshold (1 dB and 10 °) or supra-threshold (2 dB and 30 °) for both listener groups. *Right panel*, Mean peak-to-valley difference in AIM output across the three vowel response categories as characterized by the difference between the summed model output of four channels in the F2 regions of each vowel and the summed output of four channels in adjacent spectral valleys.

CONCLUSION

Impaired frequency resolution results in smearing of spectral information, potentially interfering with speech understanding, particularly in noise. The results presented here suggest that a combination of cues from component levels and phases of harmonic complexes may be exploited to provide more definition in the cochlear representation of formant structure. Although there were large individual differences in performance based on both amplitude and phase contrasts, for both normal-hearing and hearing-impaired listeners, all listeners showed an ability to improve identification performance with increasing stimulus contrast along both dimensions. Moreover, cues from amplitude and phase contrasts may combine to increase overall spectral contrast in both normal and impaired auditory systems.

The pattern of results observed for the three types of stimuli can be explained by a model of auditory processing at the level of the cochlea that accounts for the temporal response as well as the place/frequency response. Models of processing that simulate only the spectral response of the auditory periphery and discard temporal (phase) information (such as excitation pattern models) cannot account for listeners' ability to identify these vowel-like sounds coded by phase shifts. However, representations developed from a spectro-temporal model showed changes that corresponded with patterns in the behavioral identification data. In addition, the added enhancement observed behaviorally when combining amplitude and phase cues was encoded in the spectro-temporal excitation patterns as "formant" peak amplitude increments at the output of the model.

The results of this study support the idea that both NH and HI listeners can improve their identification of vowel-like sounds by combining appropriate phase and amplitude information to code formant regions. Obviously, the stimuli used here are much simpler than natural speech stimuli, but this study provides a

"proof of concept" that identification can be improved for a limited set of complex vowel-like sounds by combining two different cues to formant identity, even though those cues individually may not be useful to a listener. Further evaluation of phase enhancements of more realistic speech signals under challenging listening conditions may demonstrate whether some similar combination of spectral and temporal cues may result in signal processing improvements to speech recognition.

ACKNOWLEDGMENTS

This research was supported by grant no. R01 DC 00626 [PI: Leek] from the NIDCD. Diedesch was supported by T35 DC008764 [PI: Leek], also from the NIDCD. Support was also provided by the Department of Veterans Affairs Rehabilitation Research & Development Service [Career Development grants C6116W (Molis) and C4963W (Gallun), and Senior Research Career Scientist award C4042L (Leek)]. The work was supported with resources and the use of facilities at the Portland VA Medical Center. The contents of this article do not represent the views of the Department of Veterans Affairs or the United States Government.

REFERENCES

- ALCÁNTARA JI, MOORE BCJ (1995) The identification of vowel-like harmonic complexes: effects of component phase, level, and fundamental frequency. *J Acoust Soc Am* 97:3813–3824
- ALCÁNTARA JI, HOLUBE I, MOORE BCJ (1996) Effects of phase and level on vowel identification: data and predictions based on a nonlinear basilar-membrane model. *J Acoust Soc Am* 100:2382–2392
- AMERICAN NATIONAL STANDARDS INSTITUTE (2004) Specifications for audiometers (ANSI S3.6-2004). American National Standards Institute, New York
- BAER T, MOORE BCJ, GATEHOUSE S (1993) Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times. *J Rehab Res Dev* 30:49–72
- BLEECK S, IVES T, PATTERSON RD (2004) Aim-mat: the auditory image model in MATLAB. *Acta Acust United with Acustica* 90:781–787

- DiGIOVANNI JJ, NELSON PB, SCHLAUCH RS (2005) A psychophysical evaluation of spectral enhancement. *J Speech Lang Hear Res* 48:1121–1135
- DREISBACH LE, LEEK MR, LENTZ JJ (2005) Perception of spectral contrast by hearing-impaired listeners. *J Speech Lang Hear Res* 48:910–921
- GIGUÈRE CD, SMOORENBURG GF (1998) Computational modeling of outer hair cell damage: implications for hearing aid signal processing. In: Dau T, Kollmeier B, Hohmann V (eds) *Psychophysics, physiology and models of hearing*. World Scientific, Singapore, p 155–164
- GLASBERG BR, MOORE BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47:103–138
- GLASBERG BR, MOORE BCJ (2002) A model of loudness applicable to time-varying sounds. *J Aud Eng Soc* 50:331–342
- GOCKEL H, MOORE BCJ, PATTERSON RD (2003) Louder sounds can produce less forward masking: effects of component phase in complex tones. *J Acoust Soc Am* 114:978–990
- LEEK MR, SUMMERS V (1996) Reduced frequency resolution and the preservation of spectral contrast in noise. *J Acoust Soc Am* 100:1796–1806
- LEEK MR, DORMAN MF, SUMMERFIELD Q (1987) Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 81:148–154
- MILLER RL, CALHOUN BM, YOUNG ED (1999) Contrast enhancement improves the representation of /eh/-like vowels in the hearing-impaired auditory nerve. *J Acoust Soc Am* 106:2693–2708
- MOORE BCJ (2007) *Cochlear hearing loss: physiological, psychological and technical issues*, 2nd edn. Wiley, Chichester
- MOORE BC, GLASBERG BR (2004) A revised model of loudness perception applied to cochlear hearing loss. *Hear Res* 188:70–88
- PATTERSON RD (1994) The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96:1409–1418
- PATTERSON RD (2000) Auditory images: how complex sounds are represented in the auditory system. *J Acoust Soc Jap(E)* 21:183–190
- PATTERSON RD, ROBINSON K, HOLDSWORTH J, McKEOWN D, ZHANG C, ALLERHAND M (1992) Complex sounds and auditory images. In: Cazals Y, Horner K, Demany L (eds) *Auditory physiology and perception*. Pergamon, Oxford, p 429–443
- PATTERSON RD, ALLERHAND MH, GIGUÈRE C (1995) Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am* 98:1890–1894
- PETERSON GE, BARNEY HL (1952) Control methods used in the study of vowels. *J Acoust Soc Am* 24:175–184
- SIMPSON AM, MOORE BCJ, GLASBERG BR (1990) Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Otolaryngol Suppl* 469:101–107
- TRAUNMÜLLER H (1987) Phase vowels. In: Schouten MEH (ed) *The psychophysics of speech perception*. Martinus Nijhoff, Dordrecht, p 337–384