SELECTIVE WEIGHTING OF ACOUSTIC INFORMATION IN DISCRIMINATION OF SPEECH AND NON-SPEECH SOUNDS

Peter J. Bailey, Nicholas I. Hill* and Philip Hodgson§*
**Department of Psychology, University of York, York YO1 5DD, U.K.*
§Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, U.K.

ABSTRACT

The first experiment demonstrated that auditory discrimination of complex tones was enhanced when the distinctive information was in an expected frequency region. The second experiment sought to establish whether listeners discriminating speech sounds gave greater perceptual weight to frequency regions in which distinctive information was present.

INTRODUCTION

It is well established that listeners are more successful at detecting weak tonal signals presented in noise when the signal energy is in an expected frequency region, compared with detection performance when the signal energy is in an unexpected frequency region [e.g. 1].

The fact that observers engaged in detection tasks can apparently modulate the contribution that acoustic energy in a given spectral region makes to the resultant percept raises the possibility that such differential weighting of information may be employed in more general listening situations. For example, continuous speech typically provides unfolding information about upcoming segments which could be exploited by the auditory system to optimise recognition performance by directing attention to critical spectro-temporal regions.

However, it is unclear to what extent the results obtained using detection tasks extend to the processing of supra-threshold stimuli. To this end we have recently developed a novel analogue of the probe-signal method which can be used to explore the distribution of attention in auditory *discrimination* tasks.

The probe-signal method [2] measures the detectability of sinusoidal probe signals of various frequencies in circumstances where observers' attention is directed to a frequently-presented fixed-frequency primary signal. Detection rates for probe signals typically decrease as the frequency separation of probes and primary signal increases, a result consistent with the idea that observers attend principally to energy in the vicinity of the primary signal.

EXPERIMENT 1

The first experiment used our variant of this procedure to investigate the effect of spectral attention on listeners' ability to discriminate between multi-component complex tones distinguished by an amplitude increment in a single component.

Stimuli and Equipment

The standard stimulus was a complex tone comprising seven logarithmically-spaced components ranging in frequency from 200 Hz to 1211 Hz. All components were 200 ms in duration including 10 ms raised-cosine onset and offset ramps. Components were gated synchronously, and were each presented monaurally at 40 dB SPL.

Procedure

Listeners were required to discriminate between the standard and standard-plus-signal, where the signal was an increment in the amplitude of one of components three through six. Discrimination performance was measured using a four-interval, two-alternative, forced-choice procedure with the signal added to the standard in either interval two or three. Feedback was provided.

The effect of attention was examined as follows. In Condition 1, attention was directed towards component 3 by presenting supra-threshold (primary) increments to that component on two-thirds of trials. On the remaining trials probe-increments were applied to one of the components three through six. Condition 2 was identical except that primary-increments were applied to component 6. In both conditions the component to be incremented was selected randomly on each trial.

Probe-increment magnitudes corresponding to 79% correct detection were determined separately for each component and observer. The level of the primary increments was set at 8 dB above that of the corresponding probe.

Each subject took part in four sessions, two per condition. Each session was introduced by a practice block in which only the primary was presented. Data were collected for four subjects (2 male, 2 female) with the order of the conditions counterbalanced across subjects.

Results

The results are shown in Figure 1. Each panel gives the data for an individual subject.

Discussion

For all four observers, discrimination performance was best when increments were applied to the expected component despite the fact that all increments were nominally equally discriminable. Specifically, discrimination performance for increments applied to component six was significantly better when observers were encouraged to attend to that component, than when their attention was directed towards component three. The results demonstrate a clear effect of context on discrimination performance for the probe targets, and are consistent with the hypothesis that observers were attending preferentially to the cued component.

EXPERIMENT 2

The preceding demonstration of differential weighting of information in a discrimination task encourages the possibility that similar perceptual strategies may be employed in more general listening situations.

In an attempt to demonstrate spectrally-focussed listening during speech perception, Scharf *et al.* [3] examined the effect on performance of adding a weak tonal cue to a synthesised [da] syllable in a [da]-[ga] discrimination task. The syllables were distinguished primarily by the direction of the third formant transition. Both syllables were presented in band-limited noise centered on the frequency of the third formant. In the absence of a tonal cue, identification performance was approximately 55%. Adding a tonal cue to the [da] syllable mid-way in frequency between the first and second formants had little effect, with performance still below 60%. However, when the tonal cue was centered on the frequency of the third formant, identification performance exceeded 70%.

This may indicate that listeners were attending primarily to the third-formant region, but an alternative possibility is that the differing effectiveness of the two tonal cues may have resulted from the differential effects of the added energy on the phonetic quality of the [da] percept.

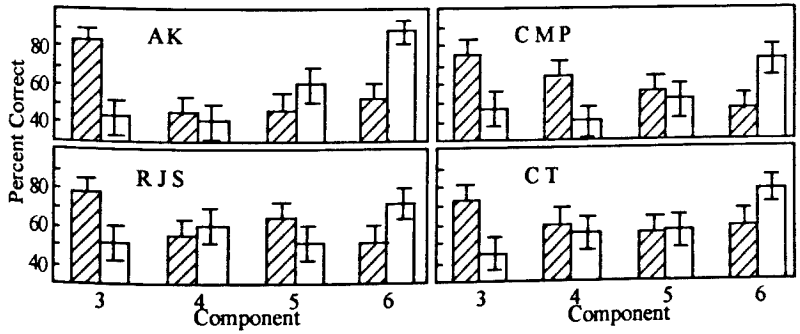

Figure 1. Discrimination performance for targets comprising threshold increments to the 3rd, 4th, 5th and 6th components respectively. Hatched columns denote performance when attention was focused on the 3rd component, open colums that when attention was focused on the 6th component. Error bars denote 95% confidence intervals.

The following experiment attempted to distinguish between these two possibilitites. Consonant discrimination performance was measured for two groups of listeners. For one group tonal energy was added to the [da] syllable, and for the other group the tone was added to the [ga] syllable. If the tonal energy merely serves as a cue to syllable identity and has no effect on phonetic quality, both groups of listeners should perform best when the tonal cue is most detectable. Accordingly, if listeners direct attention preferentially to the region encompassing the third formant transition, then performance should be better when the cue tone is centered on F3 than when it is centered on F2.

Stimuli and Equipment

Schematic versions of the syllables [da] and [ga] were created using a version of the Klatt synthesiser. They were 300 ms in duration. The difference between [da] and [ga] was carried only by a 40 ms third formant transition. Stimuli were presented monaurally in continuous white noise having an overall level of 50 dB SPL. The cue tones were 40 ms in duration including 10 ms raised-cosine onset and offset ramps, and when present were synchronised with the onsets of the syllables.

Subjects

The eight subjects had normal pure-tone thresholds up to 5 kHz. Subjects were assigned to one of two groups, with 2 males and 2 females in each group.

Procedure

Both groups of subjects were presented with equal numbers of [da] and [ga] syllables for identification, in a random order. For subjects in Group 1, each presentation of [da] was accompanied by the presentation of a cue-tone, and for subjects in Group 2 a cue-tone was added to each presentation of [ga]. The frequency of the cue-tone was either 1220 Hz or 2500 Hz, corresponding to the second and third formant frequencies of the [a] vowel. Cue-tones were presented in a random order with the restriction that both frequencies occured equally often.

Each subject was run on a stimulus calibration session followed by a data collection session. The calibration session was used to determine the appropriate levels for the syllables and cue-tones. The level of the syllables was set to give 71% correct identification, and the level of the cue-tones was set to give 79% correct detection when presented with the appropriate syllable.

Results.

Figure 2 shows the individual identification performance together with the 95% confidence intervals for the two groups of listeners. The open and hatched columns in each panel denote performance for the syllable presented,

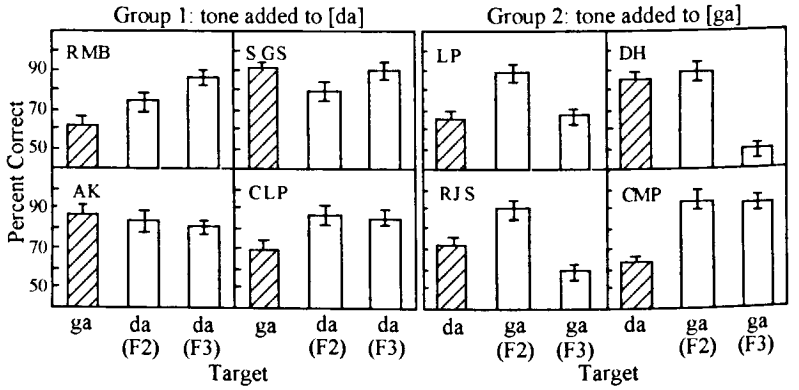

Figure 2. Percent correct identification together with 95% confidence intervals for the two groups of subjects in Experiment 2.

respectively, with and without an added tone.

Discussion

The effect of tone frequency on identification of the [da] syllable (Group 1) was significant for two of the four listeners (two-tailed test for proportion, RMB: $Z = 3.80$, $p < 0.001$; SGS: $Z = 3.84$, $p < 0.001$), with performance in both cases being better when the tone was added to the third-formant region. Three of the four listeners in Group 2 also showed a significant effect of tone frequency on syllable identification (two-tailed test for proportion, LP: $Z = -6.33$, $p < 0.001$; DH: $Z = -9.93$, $p < 0.001$; RJS: $Z = -8.78$, $p < 0.001$), but in this case performance was better when the tone was added to the second-formant region.

The hypothesis that attention was directed preferentially to the third-formant region predicts that the addition of extra distinctive energy in that region should have increased performance for both groups of subjects. The results do not support this hypothesis.

Given the absence of feedback, the fact that none of the listeners reported hearing the cue tones, and the contrast between the Group 1 and Group 2 data, the most plausible explanation of the results is that addition of tonal energy influenced the phonetic quality of the syllable to which it was added. Energy added at syllable onset in the third-formant region tended to increase the accuracy of identification of the consonant with alveolar place of articulation, whereas energy added in the second-formant region tended to increase the accuracy with which consonants with velar place of articulation were identified.

This is broadly consistent with the results of Stevens and Blumstein [4], who have argued that an important cue for the place of articulation of initial stop consonants is the distribution of energy across frequency in the gross amplitude spectrum sampled at consonant release. Their data suggest that the probability of alveolar responses should increase with increased high-frequency spectral energy at onset, and the probability of velar responses should increase with increased mid-frequency energy at onset, a pattern which is evident in our data. The pattern of consonant identification performance shown by Scharf et al. [3] is also consistent with this account.

The data were characterised by significant between-subject variability, which may reflect differences between subjects in the pattern of cue usage in identification of place of articulation [5].

CONCLUSION

The results of the first experiment demonstrated strategic weighting of spectral information in a discrimination task.

The results of the second experiment suggested that when testing the intuitively-plausible hypothesis that speech recognition accuracy is maximised by selective attention to those frequency regions in which distinctive energy is present, care is needed to dissociate attentional effects from those attributable to phonetic processing.

ACKNOWLEDGEMENT

Financial support was provided by U.K. Medical Research Council project grant No. G9306195N

REFERENCES

[1] Dai, H., Scharf, B., and Buus, S. (1991). "Effective attenuation of signals in noise under focused attention," J. Acoust. Soc. Am. vol. 89, 2837-2842.

[2] Greenberg, G. Z., and Larkin, W. D. (1968). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. vol. 44, 1513-1523.

[3] Scharf, B., Dai, H., and Miller, J. L. (1988). "The role of attention in speech perception," J. Acoust. Soc. Am., vol. 84, S158.

[4] Stevens, K. S. and Blumstein, S. E. (1978) "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am., vol. 64, 1358-1368.

[5] Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (1983) "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants", J. Acoust. Soc. Am., vol. 73, 1779-1793.