Abstract: Vowels are usually characterised by the profile of spectral amplitude across frequency. Peaks in the profile corresponding to formants provide the principal cues for identification. Such peaks can be located when stimuli are presented to either ear alone. Hence we refer to them here as "monaural" cues. Experiment 1 demonstrated that vowels can also be identified from the profile of interaural decorrelation across frequency in a noise with a flat amplitude spectrum. Peaks in the profile of interaural decorrelation can be recovered only when both ears are stimulated together. Hence we refer to them as "binaural" cues. Accuracy in identifying 2-formant vowels was not materially lower when one formant was defined binaurally and the other monaurally. Thus listeners can integrate binaural and monaural evidence of vowels. Experiment 2 demonstrated that accuracy in identifying a monaural vowel was reduced if a binaural vowel was also present, and vice versa. Thus, attention cannot be directed exclusively to either profile. Integration of the profiles is advantageous, and should maximise intelligibility over a range of signal to noise ratios when natural speech and a competing noise originate from different lateral positions.

EXPERIMENT 1: INTEGRATION OF MONAURAL AND BINAURAL INFORMATION

A low-frequency tone can be detected in a diotic noise at a particularly adverse signal-to-noise ratio (SNR) when tone and noise originate from different simulated lateral positions. Detection is facilitated, in part, because the addition of the tone to the noise reduces the interaural correlation of the noise, and listeners are highly sensitive to interaural de-correlation (1). Similar, though smaller, advantages are found for the identification of broad-band signals, including speech, in noise, where listeners can access the profile of interaural decorrelation across frequency (2). As the SNR is reduced, evidence of the spectral structure of the speech signal increases in the profile of interaural decorrelation, but decreases in the profile of spectral amplitude. Thus, intelligibility would be maximised throughout a range of SNRs if listeners could combine evidence from the two profiles. Experiment 1 asks whether such integration is possible. Experiment 2 asks whether integration is optional or obligatory.

Stimuli were constructed from 500-ms segments of diotic white noise, low-pass filtered at 4 kHz, with a spectrum level of 35 dB SPL. One-ERB wide bands (3) within the segments were modified in two ways to create rectangular approximations to formants. "Binaural" formants were created by setting the interaural correlation of a band to zero (i.e., by synthesising the band in the left and right ears from independent noise sources). "Monaural" formants were created by increasing the level of a band by 6 dB at both ears, thus creating 6 dB of "spectral contrast". Formants were centred on 250, 650, 950, and 1850 Hz. By combining pairs of formants, four 2-formant vowel-like sounds were defined: "oo" (250 and 950 Hz), "ee" (250 and 1850 Hz), "ar" (650 and 950 Hz), and "air" (650 and 1850 Hz). Note that both formants are required to specify each vowel uniquely. Stimuli were created in which both formants were monaural ("MM"), both were binaural ("BB"), the lower formant was monaural and the higher formant monaural ("MB"), and the reverse ("BM"). In addition, two sets of control stimuli contained only a single formant, either monaural ("M") or binaural ("B"). The four 2-formant conditions and the two 1-formant conditions were randomised together and presented to listeners whose task was to identify the vowel presented on each trial.

Figure 1 shows that accuracy of identification was almost as high with two binaural formants (BB) as with two monaural formants (MM). Thus, listeners can identify vowels from the profile of interaural decorrelation. In the mixed-mode conditions (MB and BM), accuracy was not materially lower than in the single-mode conditions (MM and BB). To identify the strategy used by listeners to combine monaural and binaural formants, accuracy in the mixed-mode conditions was predicted from accuracy in the 1-formant control conditions using a simple additive model. If a monaural formant was identified as vowel A on x% of trials, and a binaural formant was identified as vowel A on y% of trials, then the percentage of responses given to vowel A when the two formants were presented together was given by (x+y)/2. These predictions are plotted in Figure 1 as the filled symbols; they are all lower than the accuracy achieved in the mixed-mode conditions. Thus, listeners did not identify the mixed-mode stimuli simply by averaging their estimates of the vowels defined by the individual formants. Instead, it is likely that they based their responses on the composite pattern defined by the monaural and binaural formants together.

1975
FIGURE 1. Results of Experiment 1. Open symbols: accuracy of identification in the 2-formant conditions. Filled symbols: predictions for mixed-mode performance from performance in 1-formant control conditions. The errorbars plot 95%-confidence intervals calculated using the binomial distribution. Each point is based on 192 trials.

EXPERIMENT 2: SEGREGATION OF MONOAURAL AND BINAURAL INFORMATION

Experiment 2 asked whether the integration seen in Experiment 1 is optional or obligatory. Binaural (BB) vowels have a tonal timbre and are lateralised, whereas monaural (MM) vowels have a noisy timbre and are heard on the mid-line. Thus, listeners can be instructed to attend selectively to either type of vowel. In the main conditions of Experiment 2, each stimulus contained 2 monaural formants and 2 binaural formants, defining two different vowels—one monaural, the other binaural. In separate conditions, listeners were instructed to report either the binaural vowel or the monaural vowel. Accuracy was measured as a function of the spectral contrast of the monaural formants (0, 3, 6, 9, or 12 dB).

Figure 2A shows that accuracy in identifying the binaural vowel declined as the spectral contrast of the monaural formants was increased. Figure 2B shows that accuracy in identifying the monaural vowel was lower when a binaural vowel was present (open symbols) than when no binaural vowel was present (filled symbols). Thus, listeners could not ignore the profile of spectral amplitude when attempting to attend to the profile of interaural decorrelation, or vice versa: some degree of integration is obligatory.

A unifying account of the results of Experiments 1 and 2 is that there is a strong tendency to integrate the two profiles, and that peaks in the combined profile are only loosely flagged as to their origin. As argued in the Introduction, such integration maximises the intelligibility of natural speech as the signal-to-noise ratio varies.

REFERENCES