Poster presented at the BSA short papers meeting, Sept 27-28, 1995, Oxford, UK; a version of this will appear in the British Journal of Audiology.

A simple model of amplitude modulation detection.

L.S. Smith, CCCN/Department of Computing Science, University of Stirling, Stirling FK9 4LA, Scotland. lss@cs.stir.ac.uk

Deterioration of hearing at high frequencies leads to problems in speech interpretation in noise. One candidate for a carrier of useful information at higher frequencies is amplitude modulation (AM) found in wideband bandpassed voiced speech due to unresolved F_0 harmonics. Taking an approach based in auditory modelling, we seek to identify (and eventually characterise) voiced sounds.

Initial processing of digitised speech was Gammatone filtering (with bandwidth $C_f/9.265 + 24.7$ Hz) followed by rectification. Each channel was convolved with a function composed from the difference between two half-Gaussians (both with maximum at $t=0$, and 0 for $t<0$, so that convolution does not extend forward in time). Rapid increases (onsets) in signal strength result in a positive output, while rapid decreases (offsets) result in a negative output. An onset signal was produced by logarithmically compressing the rectified convolution output. An offset signal was produced by logarithmically compressing the inverted convolution output. The effect is that rapid increases in energy in a band produce a positive pulse in the onset signal, and rapid decreases in energy produce a pulse in the offset signal. For appropriate parameter choice, this brings out the AM in a speech signal: figure 1a shows the onset signal for the word ‘she’ (duration 0.2s). To permit rapid discovery of voicing, and to allow identification and isolation of channels taking part in concurrent AM, these pulses were turned into spikes. Two techniques were tried. Firstly, an array of leaky integrator ‘neurons’, one per channel, was used. Choosing the coupling strength appropriately each will either fire during an onset pulse, or, if the pulse is not strong enough, will have its activity decay to a low value before the next pulse occurs. Synchronisation across channels was encouraged using excitatory links between near-adjacent neurons. Figure 1b shows the neural output without these links. Figure 1c shows the output with fixed excitatory links; figure 1d uses links which switch on when pre- and post-synaptic units fire at the same time. The second technique used a ‘neuron’ tuned to detect AM signals: these fire in response to an onset pulse followed by an offset pulse. Figure 1e shows the output from these neurons.

Both techniques give lines of spikes across channels for voiced speech, particularly vowels, but not for sibilances, plosives, etc. The use of fixed excitation across channels straightens the lines (compare figures 1b and 1c), but tends to form lines which are artefacts. Dynamic links reduce these (compare figures 1c,1d). Both techniques are effective at discovering vowels in continuous speech, but the second needs less parameter tuning than the first. Rapid detection even of short vowels is possible using the line of spikes generated by an AM pulse.

Figure 1: