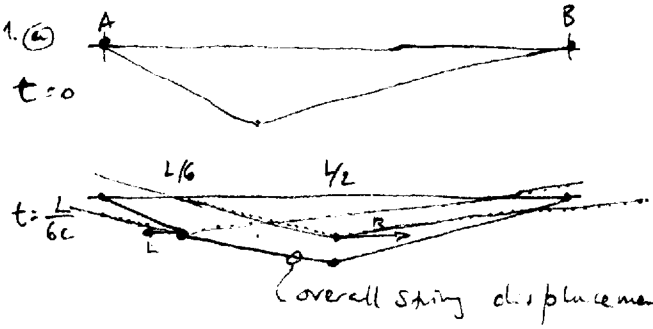
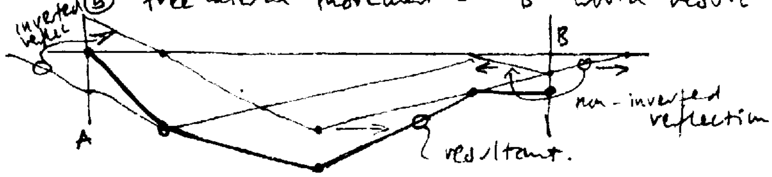
2003-03-10
dpwe@ee.columbia.edu
ELEN E6820 Midterm Solutions

1. (a)

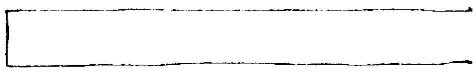


- Initial displacement is equal parts left- and right-going waves
  ⇒ images travel each way.
- Inverting reflections at fixed ends look like 180° rotated versions shifted in.
- Cancellation keeps displacement at zero at both ends

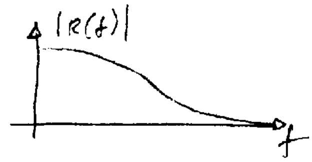
(b) free lateral movement at B would result in a non-inverting reflection:



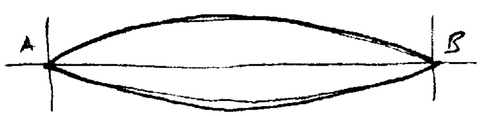
If we equate string displacement with volume velocity, fixed point at A is like a closed tube end, and free point at B is like an open end:



(c) Reflection is a low-pass filter:



⇒ higher frequency modes are attenuated relative to lowest mode at each reflection
⇒ for large T, waveform is dominated by fundamental mode
→ sinusoid standing wave:

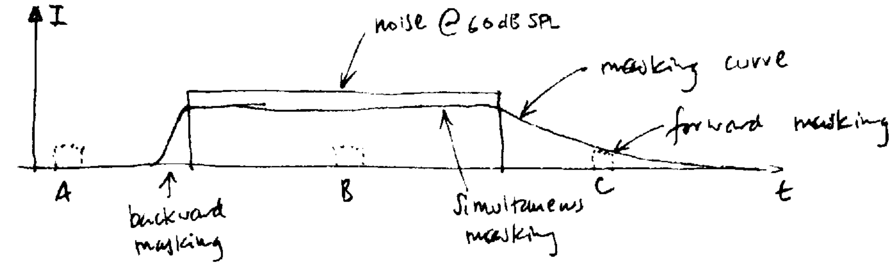


2. (a) A is a narrow-band spectrogram (analysis time window ~30ms)
B is a wide-band spectrogram (analysis time window ~5ms)
B also has a lower energy floor in its grayscale mapping
eg. A shows energy in range 30-60 dB, B shows 0-60 dB (or whatever)
hence can see more low-level energy features in B.

(b) From A, we have a pitched vowel between t = 0.2 .. 0.35.
Fundamental frequency peak ~ 220 Hz (by dividing freq. of 10th harmonic by 10). Pitch contour rises then falls.
From B, have noise between t=0.1 and 0.2, then vowel (formants at 600 Hz, 1900 Hz, 2800 Hz), then a stop, then more frication (sibilant). [Actually, the speech is "six"]

(c) A shows harmonics and hence pitch, but pitch is ignored in English ASR.
B shows the formants, which are the important features for ASR
→ B is preferable for speech recognition.

3.

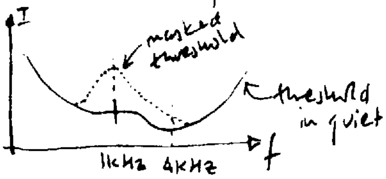


(a) Pip at A is unmasked → Just audible at around ∅ dB SPL (depending on listener, conditions)

B is simultaneously masked → inaudible until within ~ 10 dB of masker → just noticeable ~ 50dB SPL

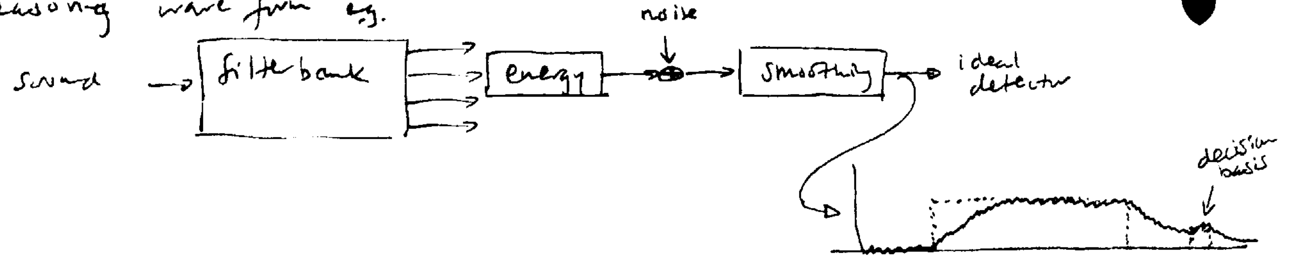
C is subject to forward masking - less than simultaneous → somewhere in between - maybe 20-30 dB SPL?

(b) Masking effect is much less at 4kHz for energy at 1kHz ∴ in a separate critical band. Ear is more sensitive at 4kHz



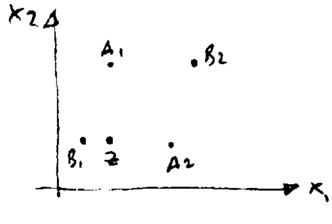
→ A noticeable @ ~ -5 dB SPL
B ~ 20 dB SPL or less
C ~ 10 dB SPL or less.

(c) · Simultaneous Masking suggests detection via energy envelope displacement.
· frequency selectivity → band-pass filtering before energy
· Forward masking suggests sluggish smoothing of energy before measuring wave form e.g.



4. (a) in $[x_1, x_2]$ space, Z is nearest A_1 → Classify to ω_A

(b) Equalizing variance in each dimension "stretches" graph vertically so 4 points lie roughly in a square:



Because transformation matrix is <u>diagonal</u> there is no rotation or shear - just stretch in x_1 and x_2

Now Z is closest B_1 → classify to ω_B

(c) Alternative approaches: Decision boundaries from · linear - (no perfect fit)
· Gauss models
· Neural network etc.

· Could try k-nearest neighbors e.g. find 3 nearest neighbors and vote.

Basically, problem is nearly impossible to decide with so little training data and apparently highly overlapped classes.