* Reference: Chin Hui Lee, Ken Stevens, Jim Glass
* Diagram: hierarchical recognition, lm/feature detection module
* Metrics
* Content (as suggested by Stefanie and Elizabeth): phonetic process instead of phonological
* Language

a.  Problem statement

b.   Previous work (included references from IEEE/ACM/etc.)

conventional – fixed-length frame, spectral feature, stochastic

old knowledge-driven approach: not stochastic, no feedback

new knowledge-driven approach: ASAT (Lee), probabilistic (Glass),

c.    Proposed technical approach and relation to previous work

**d.    Block diagram / architecture of the system you are building**

**e.    Metrics (as appropriate)**

f.     Expected outputs

g.    Other goals: publication, demonstrating a system, etc.

During the fall semester, I expect to complete and fully test the landmark-specific programs, apply several machine methods to the available data, and start extending the data structure to other acoustic cues; in the spring, I will aim at completing the software package for all acoustic cues, improve its usability, and test the performance of various probabilistic models.

I'm not sure we will get to develop new models that describe this process by the spring... or if you are referring to maching learning/detection methods for finding landmarks and/or acoustic cues, then I think that should be specified.

 A possible extension of the project would be integrating my programs with existing cue detection programs and trying to build an actual speech recognizer.

This part is also a bit uncertain, since it will involve quite a bit of programming in signal processing, and I am not sure if I/my former students/other students will be able to do this by the spring. As another extension to the stuff we are doing, it would be possible to work on the "Matcher" part of the system, where the detected/labeled landmarks+distinctive features would be "matched" to entries in the lexicon (=words). If you would like to, and Stefanie agrees is a suitable direction, that could be mentioned in the proposal, but again, I'm not sure about this.

I will also be responsible to maintain a human-readable documentation for the comparison results of the data for the purpose of direct inspection by researchers at the lab.

This part, of course, would be completely important and invaluable!

1)   In several places you mention that we will be drawing on knowledge expressed as phonological rules.  In this discussion we need to draw a firm distinction between phonological rules, on the one hand, and phonetic processes, on the other, and Elizabeth and I haven't spent much time discussing this with you.  So here is a brief introduction to this distinction; we can talk about it more when we meet next on Friday at 10 am.  
  
Phonological rules are rules that change one abstract symbol into another, i.e. they change one or more feature values of a phoneme.  As a result, that sound is now produced exactly as if it had the new feature set from the beginning---there is nothing left of the original feature value.  For example, when English put together the two morphemes in- (meaning something like 'not') and -possible, to make 'impossible', over time the feature specification for the nasal became [labial], specifying an /m/, instead of [alveolar], which specifies an /n/.  This change is reflected in the spelling (although the outcome of feature-changing rules are not always reflected in the spelling---consider, for example, the nasal sound in 'ink'---is it an /ng/?)  
  
In contrast, phonetic processes reflect things like articulatory overlap and articulatory weakening, and the results are usually only a partial change.  When a /k/ is weakened so that it looks a bit more like a fricative, it often still has a burst (resulting from release of pressure build up behind the partial constriction) but also has some fricative noise during the closure.  Or, when a coda nasal spreads its nasality into the preceding vowel, as in the 'can' of 'I can come', it can spread a little bit or a lot or even completely, but the 'intended' sound, the one that specifies the word in the mental lexicon, is still a non-nasal vowel.  
  
I hope these examples make the distinction clear to you---and that it now makes sense to you that most of the knowledge that we will be incorporating into the system is knowledge about phonetic processes that vary in their 'strength' with context---and not phonological rules.  See if you feel comfortable changing 'phonological rules' to 'phonetic processes' in your proposal.  
  
2)  We'd like you to get in the habit of using **references** in your writing when you draw on work by earlier investigators.  In this proposal you might want to include Chin-Hui Lee's papers when you talk about HMM-based approaches, and Ken Stevens' 2002 paper when you talk about the new cue-based approach, and include these papers in a Bibliography at the end.  As time goes on you will have many more references to include.  (I recently came across an early paper by Chin-Hui Lee in Speech Communication 1994, which focusses on statistically based models and is the one you probably should include, if you have time to find it online.)

🡪 Jim Glass  
  
3)  I'd like to echo Elizabeth's advice to tone down the scope of your ambition a bit---it is better to promise less and get more done than vice versa.  How about if you say something like you'll be working on other cues and on integrating the system, rather than that you plan to get all of that done by May.  I think we will all be glad if we haven't promised to dot every i and cross every t by then.  (But we very much like your ambitious determination to tackle the whole problem!  These are excellent goals, but not good promises to make.)