M.Sc. LST
Speech Science

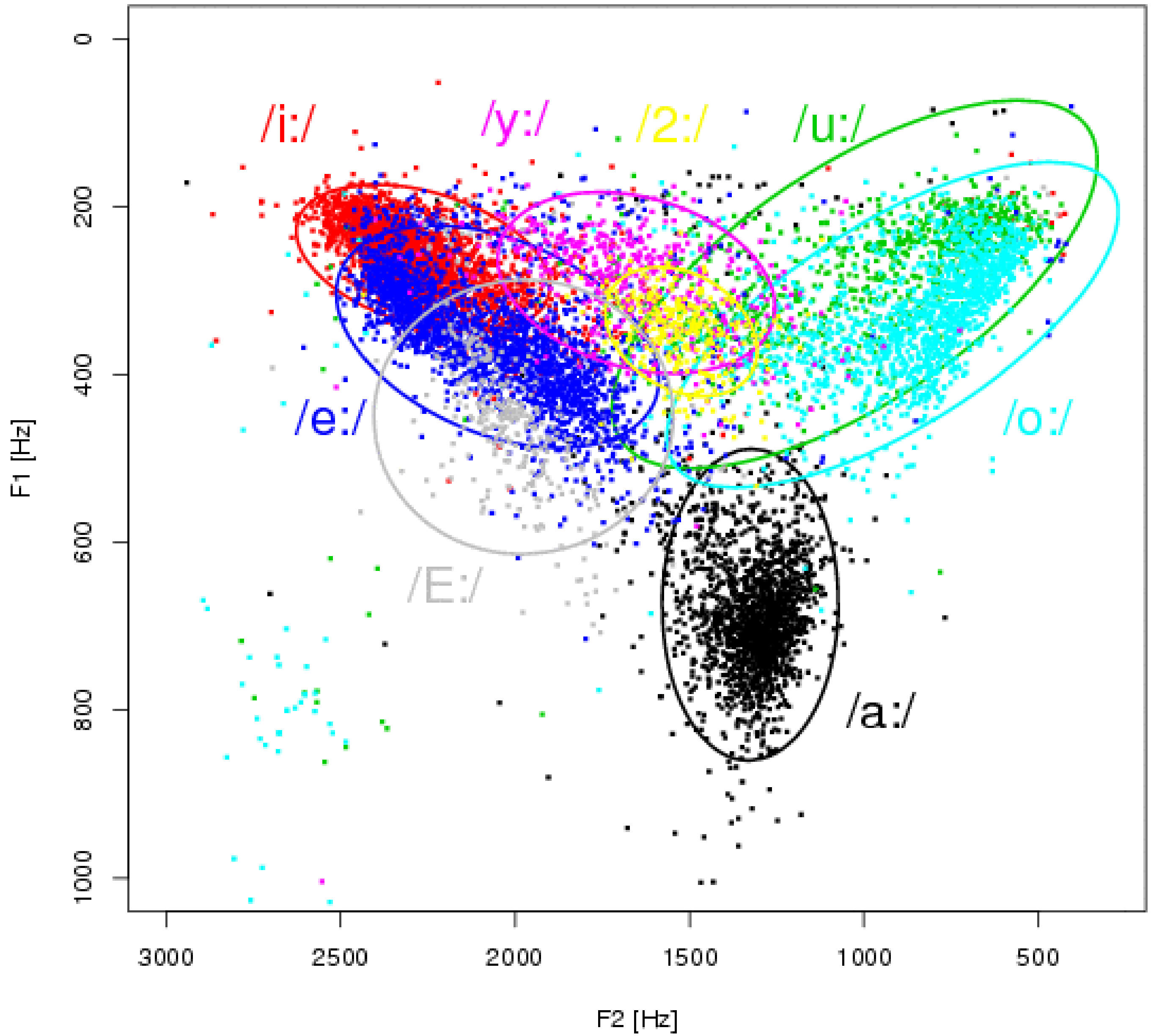
Theories and Models of Speech Perception

Feb 1, 2024

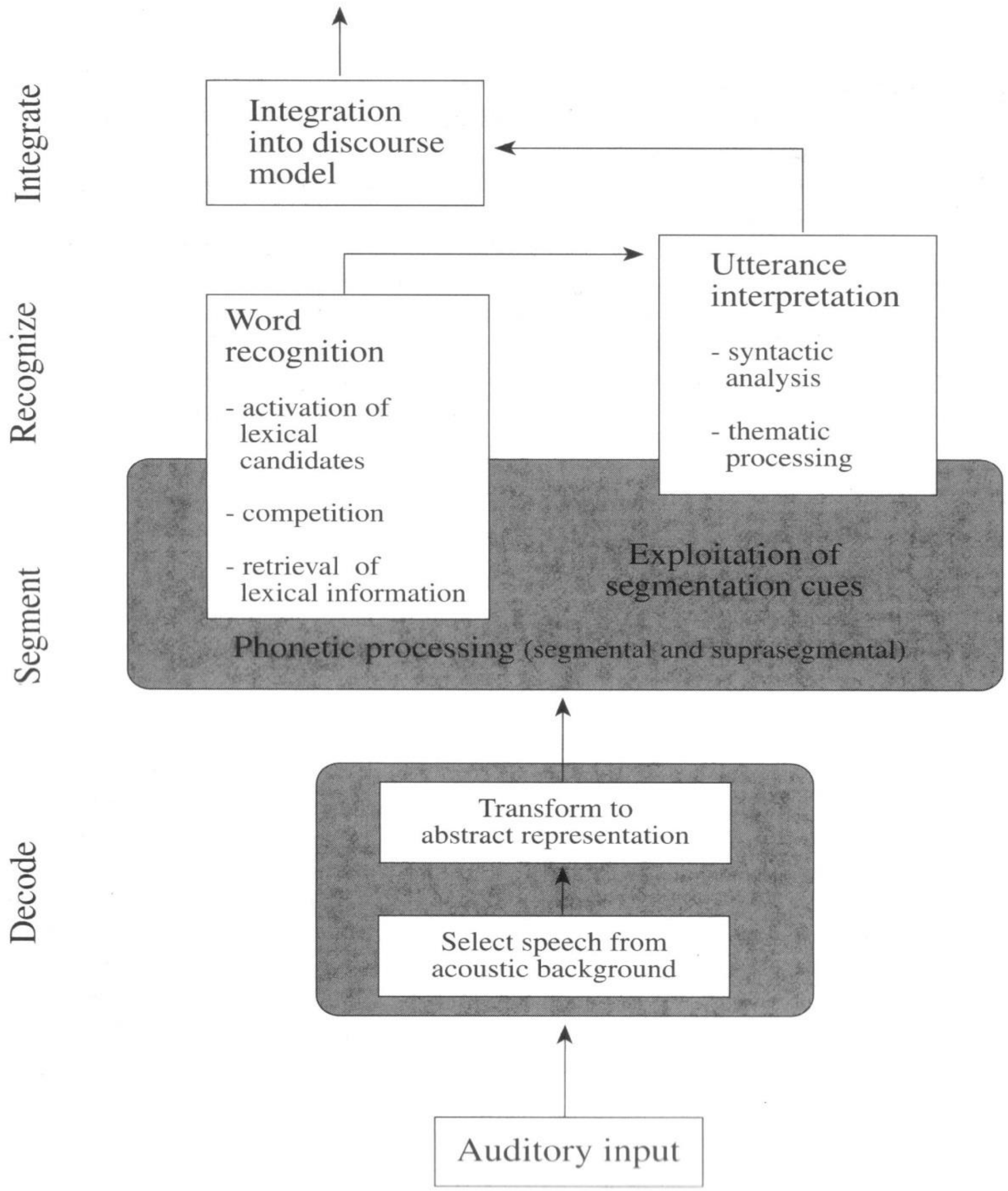
Bernd Möbius & Omnia Ibrahim

Language Science and Technology
Saarland University

Variability



A blueprint of the listener



[Cutler & Clifton 1999, p124]

Components of the blueprint

- Speech decoding: distinguish speech from other auditory input
- Segmentation of continuous signal in constituent parts
  - incremental, partially parallel processing
  - higher-level (e.g. word) processing starts before segmentation is complete
- Lexical activation: recognition of spoken words
  - activation of multiple word candidates → competition
  - relevant information: segm., suprasegm.; full/partial match?
- Morphology and word semantics from lexicon
- Syntactic relations and thematic roles
  - restriction of search space by prosody?
- Architecture of "listener"
  - degree of interactions?

Speech perception

- What are the objects of speech perception?
  - discrete segments; phone-based, syllable-based?
  - motor commands?
  - articulatory gestures?
  - vocal tract constrictions or geometries?
  - acoustic sound targets?
  - perceptually defined speech sound targets?

Speech perception

- Major theories
  - Motor Theory
  - Direct Realist Theory
  - Auditory Enhancement Theory
  - H&H Theory
  - Quantal Theory
  - (Connectionist models)
  - (Exemplar Theory)
- Phonetic "consensus model" of speech perception

Motor Theory

- first proposed in 1950s; last modified 1985 [Liberman et al. 1967, Liberman & Mattingly 1985]
- objects of perception: invariant motor gestures intended by speaker
- perceptual invariance despite vast acoustic variability
- perception relies on production, not on acoustics
- consonants are produced and perceived categorically
- vowels are produced and perceived continuously
- speech is special: phonetic module responsible for both production and perception of speech

Direct Realist Theory

- first proposed in 1980s, based on general perception theories [Fowler 1986]
- strongly related to Motor Theory
- objects of perception: discrete articulatory gestures executed by speaker
- variability arises from gestural overlap → variable coarticulation
- perceptual invariance relies on auditory separation and recoverage of gestures
- no special phonetic module
- speech perception follows general perceptual principles

Auditory enhancement

- proposed in late 1980s [Diehl & Kluender 1989]
- listeners are particularly sensitive to auditory qualities of phonetic segments (not to articulatory gestures)
- universal tendencies in sound systems of languages originate from general auditory capabilities of human listeners
- articulatory gestures are not determined predominantly by physics and physiology
- articulatory co-variation is not random but serves common goal
- gestures co-vary to jointly support certain auditory effects
- speaker and listener oriented principles
- phonetic categorization follows general auditory mechanisms
- phonetic categories are natural auditory classes, but language-specific and must be learned

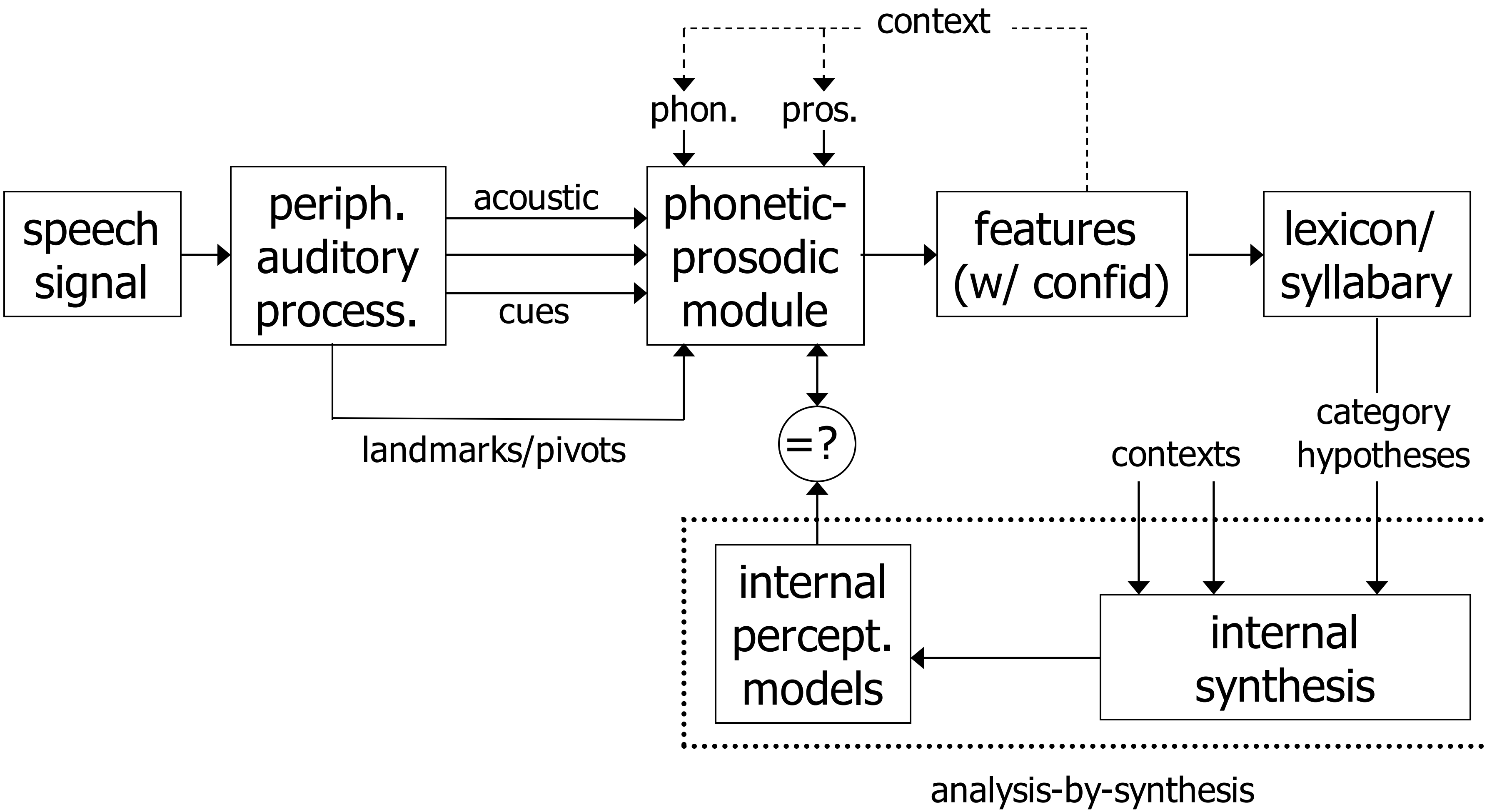
H&H Theory

- proposed in late 1980s [Lindblom 1990]
- no invariance in articulation and acoustics
- adaptive balance between hypo- and hyperarticluation
  - hypo-articulation: economy principle, principle of least effort $\rightarrow$ target undershoot, reduction
  - hyper-articulation: help listeners extract contrasts in adverse conditions or insufficient context
- encode maximum information in signal with minimal articulatory effort
- structure of speech sound inventories relies on adaptive dispersion: less vowel variability in languages with large vowel inventories

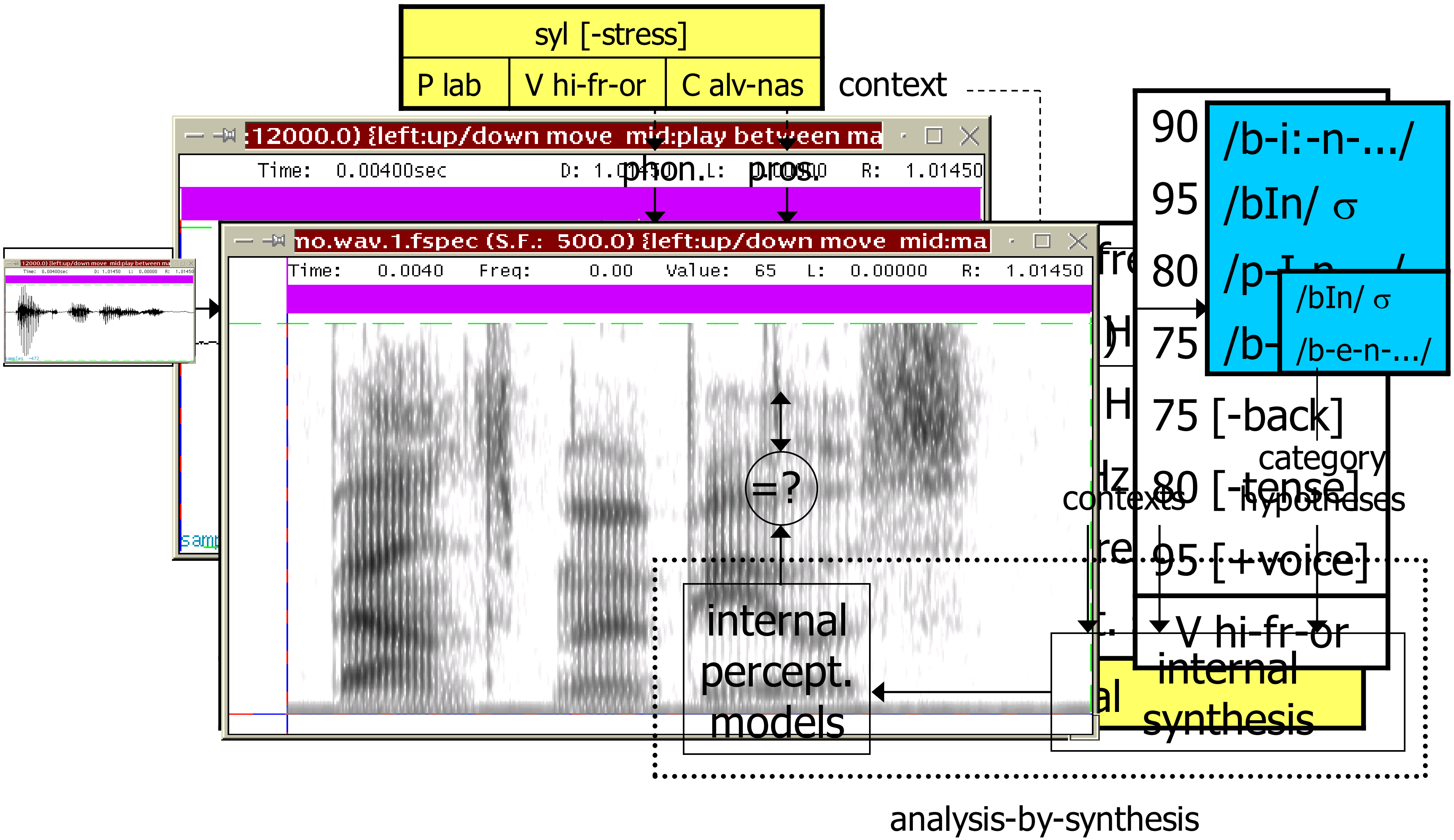
Quantal Theory

- first proposed in 1970s [Stevens 1972, 1989]
- non-linear relations between
  - articulatory space and acoustic space
  - acoustic space and auditory-perceptual space (e.g., CP)
- invariance based on non-linear relations
- invariance may be found in perception, acoustics, not in articulation
- structure of sound inventories relies on regions of invariance, phoneme boundaries in areas of quantal changes
- further developed into Lexical Access From Features model
- objects of perception: distinctive features, extracted from quantal space
- feature-based specification of mental lexicon

Phonetic "consensus" model



Phonetic "consensus" model



Model: Components

- acoustic feature extraction at key locations in speech signal
  [Stevens 1989, Dogil 1987]
- feature-based lexicon access [Stevens 2005]
- articulatory verification by means of analysis-by-synthesis
  [Gaskell et al. 1995, Stevens 2005]
- underspecified abstract lexicon and episodic exemplar lexicon
  [Dogil 2006, Möbius & Schütze 2006 (SFB)]

Model: Analysis

- incremental process of underspecification
  - extraction of acoustic parameters and robust features
  - considering contextual information (segmental, prosodic, syllable structure)
  - abstraction from speaker properties
  - lexicon access (words, morphemes, syllables, segments(?))

Model: Synthesis

- incremental process of specification
  - applied to each hypothesized category
  - internal synthesis
    - exploiting all available contextual information (segmental, prosodic, syllable structure; syntax, pragmatics)
    - transformation into perceptual space
    - fully specified representation (exemplars)
  - comparison of perceived exemplars with synthesized/stored exemplars

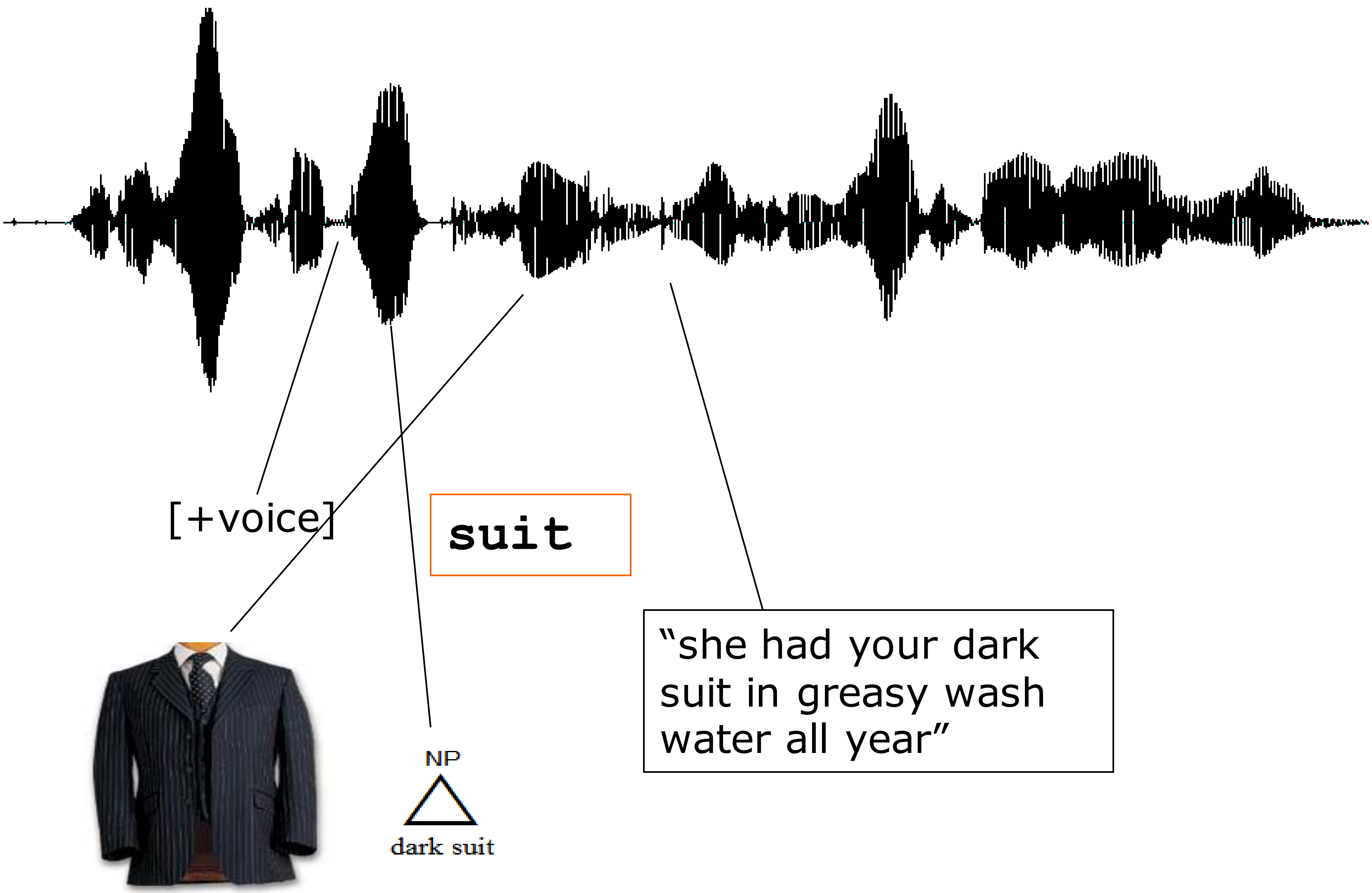
Computational model

- Why do we need a computational model?
  - requires explicit (mathematical, algorithmical) formulation
  - model-based predictions can be tested experimentally
  - interactions between assumptions can be investigated formally
  - observed behavior $\rightarrow$ model specification

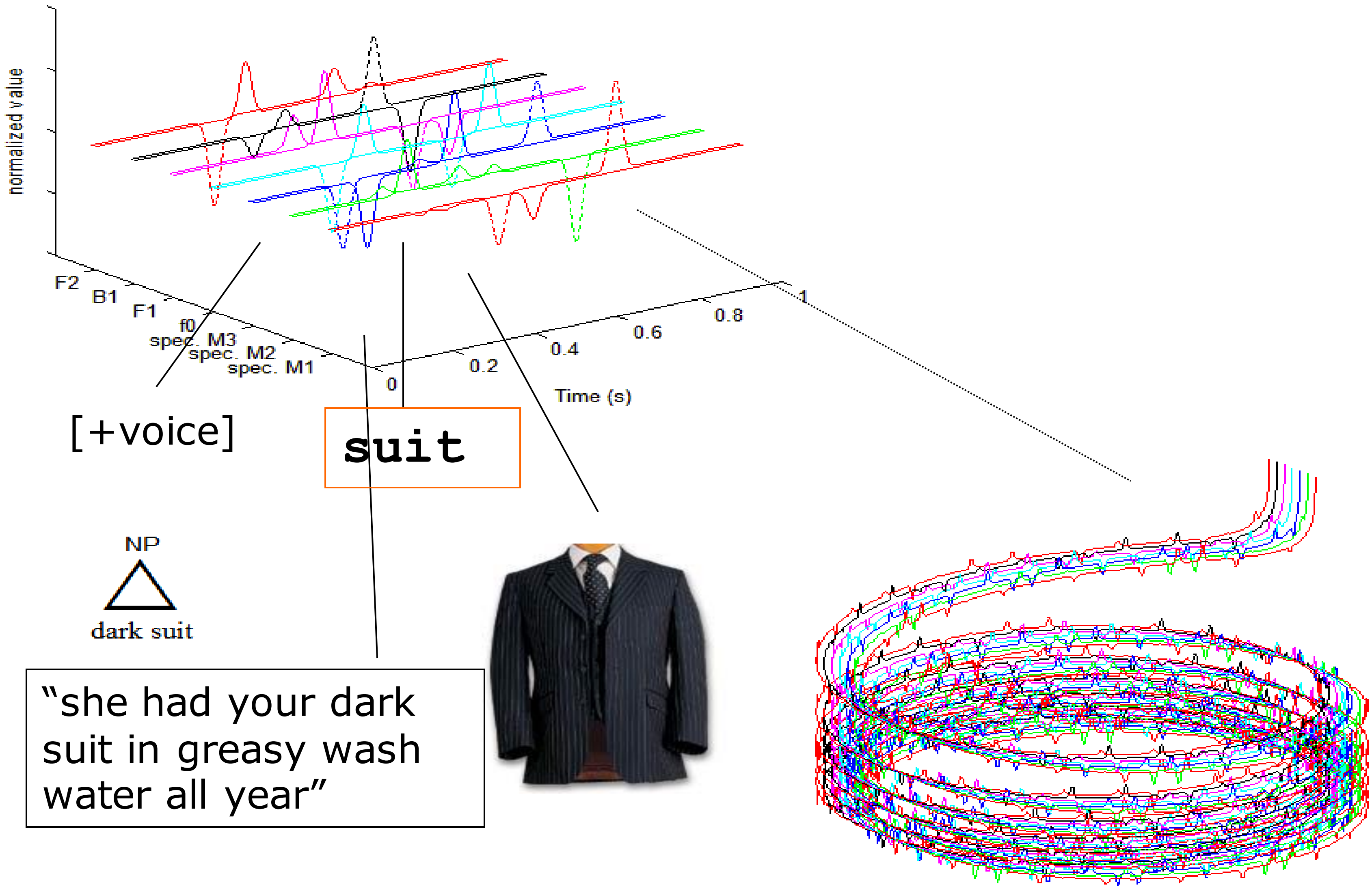
Exemplar Theory: Key assumptions

- Exemplar space: multidimensional cognitive map
  - similarity of exemplars ~ stance in this space
- Exemplars comprise detailed phonetic information (ling./paraling./extraling. dimensions) [Goldinger 1997, Pierrehumbert 2001, 2003]

Exemplar Theory: Key assumptions



Exemplar Theory: Key assumptions



Exemplar Theory: Key assumptions

- Exemplar space: multidimensional cognitive map
  - similarity of exemplars ~ stance in this space
- Exemplars comprise detailed phonetic information (ling./paraling./extraling. dimensions) [Goldinger 1997, Pierrehumbert 2001, 2003]
- Effects of frequency and recency:
  - exemplar space is updated continuously
  - memory traces decay over time

Exemplar Theory: Key assumptions

- Common levels of representation for perception and production
  - exemplars: concrete, experienced tokens
  - phonetic encoding: properties of exemplars
  - phonological encoding: category label
  - quantitative knowledge: frequency distributions

References

- Cutler A., Clifton C. (1999): Comprehending spoken language: a blueprint of the listener. In C.M. Brown, P. Hagoort (eds.), The Neurocognition of Language. Oxford Univ. Press, 123-166.
- Diehl R.L., Kluender K.R. (1989): On the objects of speech perception. Ecol. Psychology 1:121-144.
- Dogil G. (1987): Prototypical speech events and speech perception. Proc. ICPhS (Tallinn), 3:360-366.
- Dogil G. (2006): Incremental specification in context: Phonetics. SFB 732, Univ. Stuttgart. [www.uni-stuttgart.de/linguistik/sfb732/]
- Fowler C.A. (1986): An event approach to the study of speech perception from a direct-realist perspective. J.Phon., 14:3-28.
- Gaskell M.G., Hare M., Marslen-Wilson W.D. (1995): A connectionist model of phonological representation in speech perception. Cognitive Science, 19(4):407-439.
- Liberman A.M., Cooper F.S., Shankweiler D.P., Studdert-Kennedy M. (1967): Perception of the speech code. Psych. Rev., 74, 431-461.

References

- Liberman A.M., Mattingly I.G. (1985): The motor theory of speech perception revised. Cognition 21:1-36.
- Lindblom B. (1990): Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle, A. Marchal (eds.), Speech Production and Speech Modelling. Kluwer, 403-439.
- Möbius B., Schütze H. (2006): Exemplar-based speech representation. SFB 732, Univ. Stuttgart. [www.uni-stuttgart.de/linguistik/sfb732/]
- Stevens K.N. (1972): The quantal nature of speech: evidence from articulatory-acoustic data. In E.E. Davis, P.B. Denes (eds.), Human Communication: A Unifed View. 51-66.
- Stevens K.N. (1989): On the quantal nature of speech. J.Phon., 17:3-45.
- Stevens K. (2005): Features in speech perception and lexical access. In: Pisoni D.B., Remez R.E. (eds.), The Handbook of Speech Perception. Blackwell, 125-155.

Thanks!