Chapter 12

Models and theories of speech production and perception

1. THEORIES VS. MODELS
2. ISSUES: SERIAL ORDER/ DEGREES OF FREEDOM/ CONTEXT SENSITIVITY
3. MODELS: TARGET/ FEEDBACK VS. FEEDFORWARD/ DYNAMIC SYSTEMS/CONNECTIONIST (PDP)
4. LINK BETWEEN PRODUCTION ←→ PERCEPTION

Updated 2019

Model/ Theory/ Hypothesis

Model – Simplification of a system or any of its parts

Theory – underlying principles and assumptions

Hypothesis – a specific, testable prediction (typically grounded in a certain theory)

Careful:

Although last two terms are sometimes used interchangeably, a theory has been extensively tested and is generally accepted, while a hypothesis is a speculative guess that has yet to be tested

Models

- Can be manipulated in a controlled manner to test hypotheses or theories
- Can be physical, or (more commonly these days) mathematical

Model example

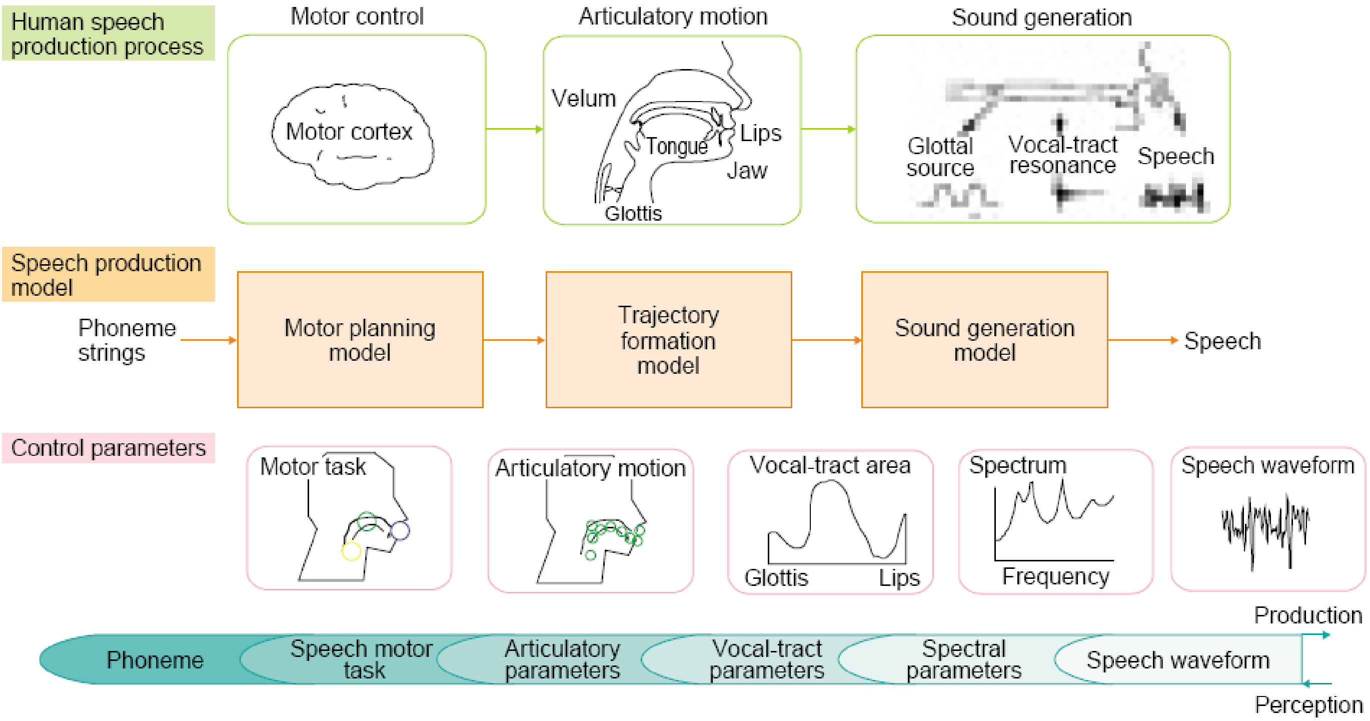


Fig. 1. Speech production process and the model.

M. Honda, NTT (2003)

The role of theory in speech science

- Theories are important!
- Our clinical tools are only as good as the theories they are based on
- A popular misconception: Theories are never ‘proven’ or ‘mis-proven’
- Instead, theories are either supported or not supported by data

Issues

1. Regulating serial order
2. Degrees of freedom
3. Context sensitivity

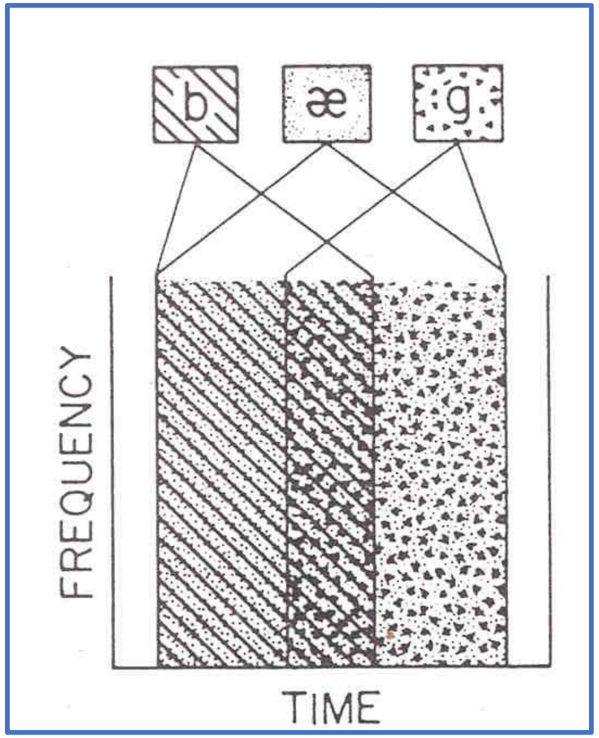
Also: Are speech goals acoustic or articulatory (or both)?

1. Regulating serial order

Coarticulation:

- Anticipatory vs. Perseverative
- Language-specific

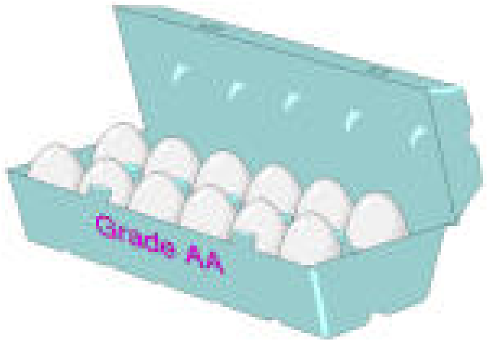
Q: How do we do it?

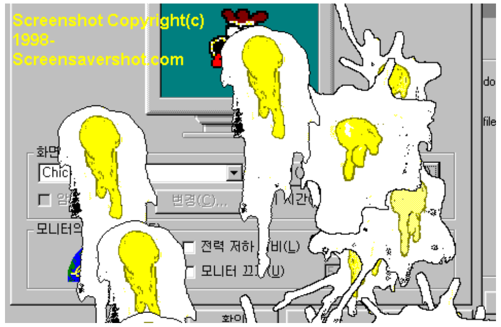


Charles Hockett (1955)

Invariant units of speech become smeared in the process of articulation

But the listener manages to recover these invariant units during perception





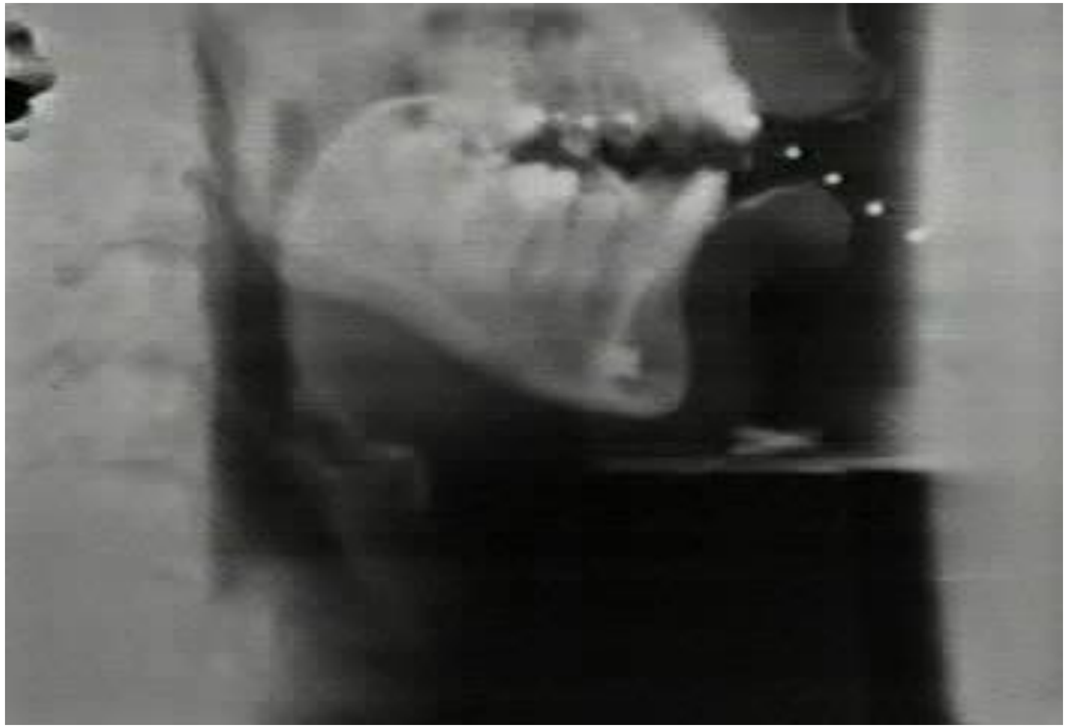
2. Controlling degrees of freedom (df)

Approx. # of muscle pairs that move the

- Tongue: 9
- Velum: 3
- Lips: 12
- Mandible: 7
- Hyoid bone: 10
- Larynx: 8
- Pharynx: 4

PLUS - muscles of the respiratory system

HOW DO WE DO IT?



- Perkell, 2003

Differing proposals…

Motor programs

Hierarchies

Functional groupings, coordinative structures

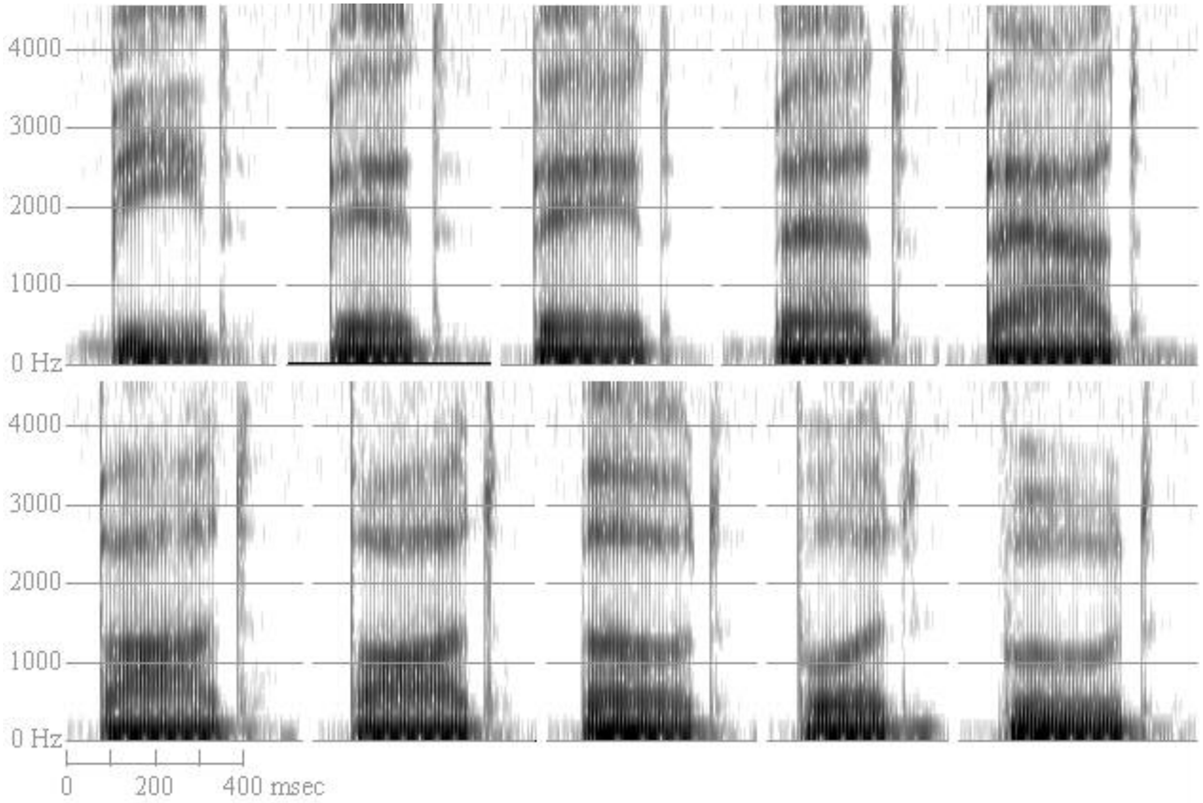
Dependency on environment, “affordances”

3. Context sensitivity

* Vowel cues are affected by their flanking consonants, and consonant cues are affected by their flanking vowels

EXAMPLE: American English vowels in /b_d/ context

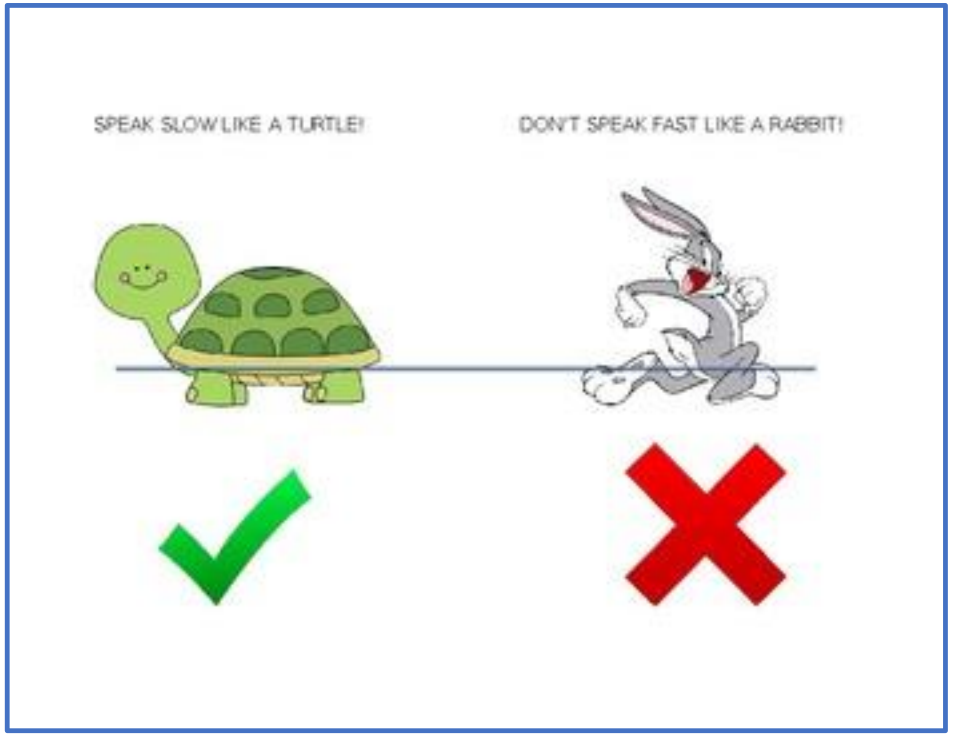
TOP ROW (front vowels): “bead bid bade bed bad”
BOTTOM ROW (back vowels) “bod bawd bode buhd booed”



Another example

Rate normalization

Same phonetic content is preserved although absolute signal differs greatly



Models of speech production

Target

Feedback vs. Feedforward

Dynamic systems/ Connectionist (PDP)

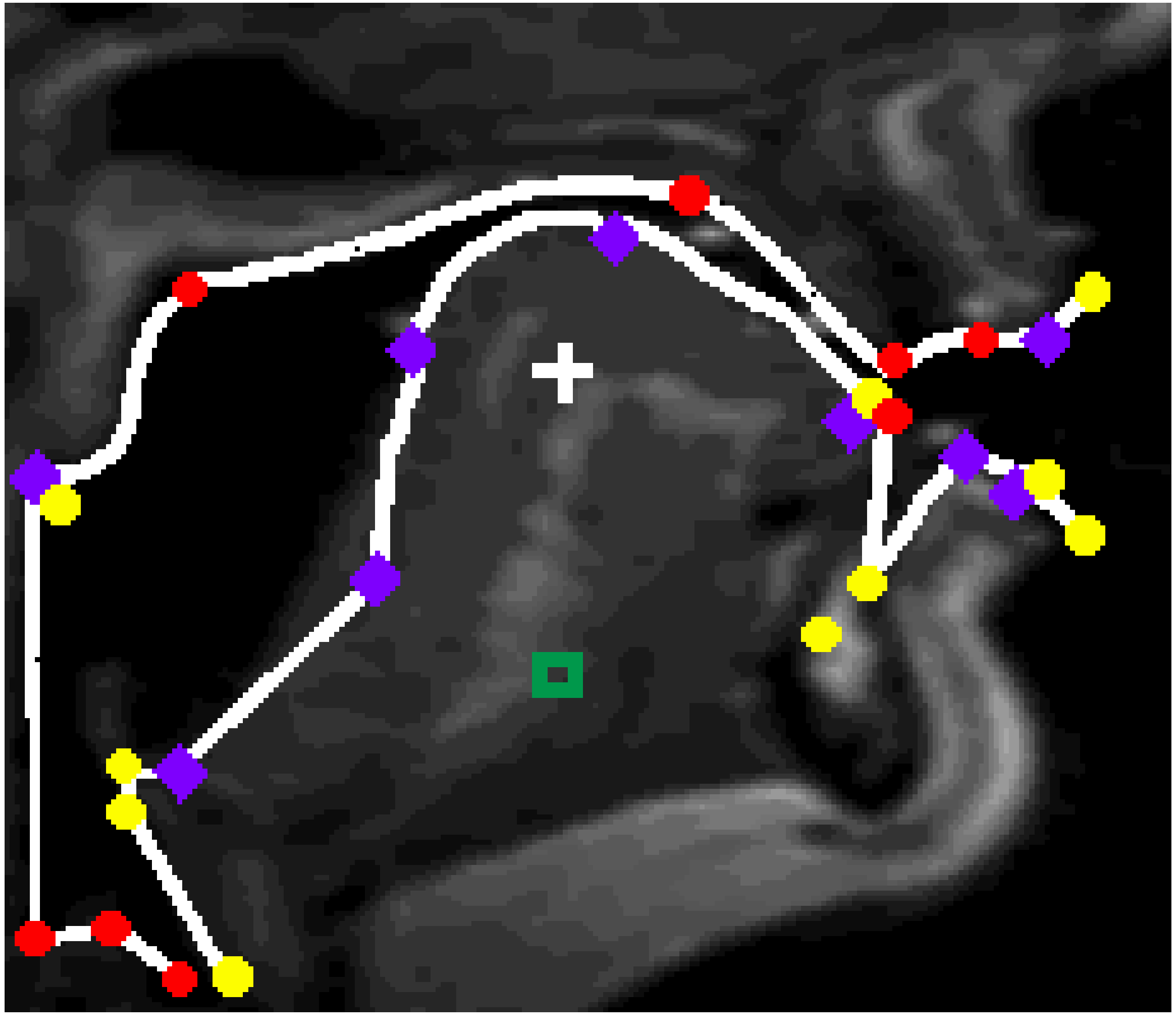
Target Models

Spatial? (e.g., “*place tongue body HERE for /k/ in the context of the preceding vowel /u/…*”)

OR

Acoustic – auditory? (e.g., “*create a silent gap with certain formant transitional characteristics…*”)

Each explanation has strengths and weaknesses

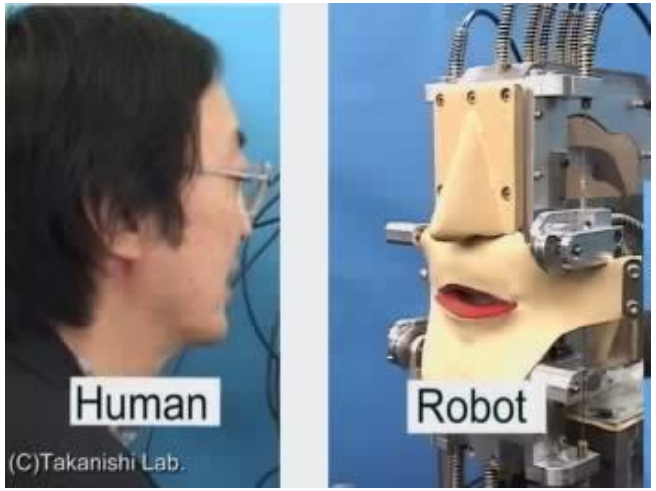


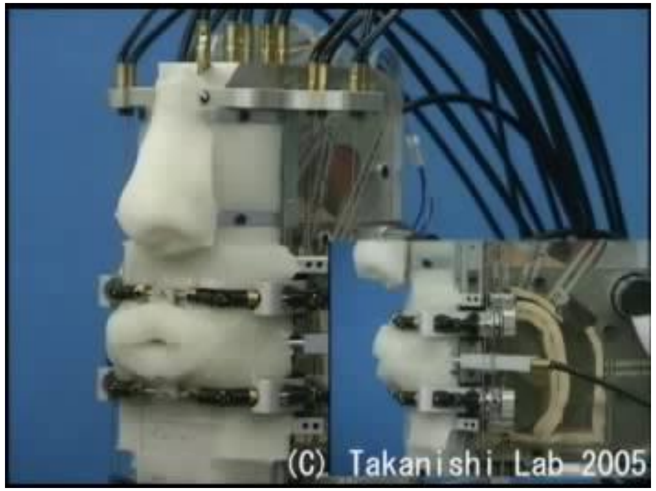
Evidence supporting the hypothesis of spatial (or articulatory) goals:

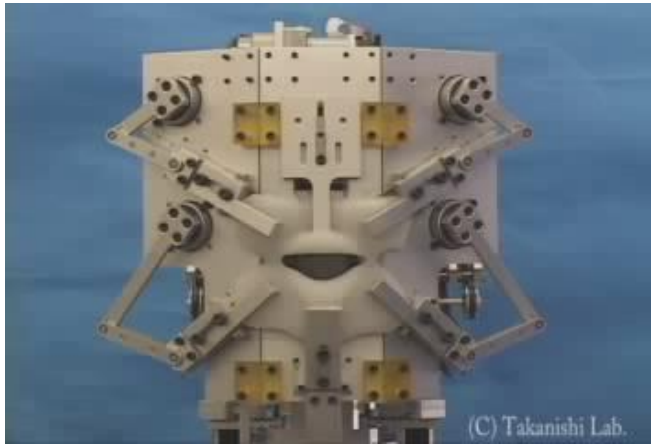
Articulatory synthesis can generate plausible vowel- and consonant- like sounds

(Mermelstein, 1972; Rubin & Goldstein, 1995)

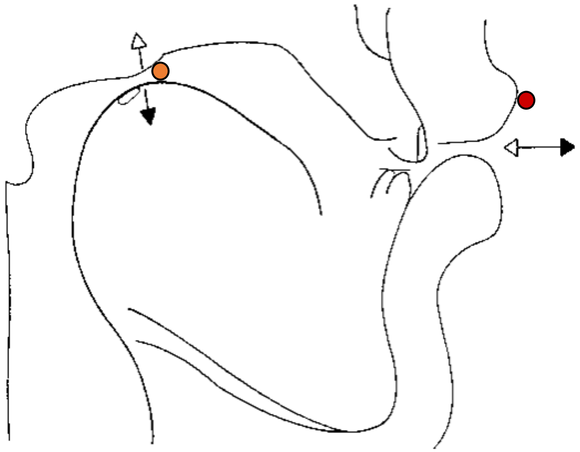
Articulatory synthesis robot



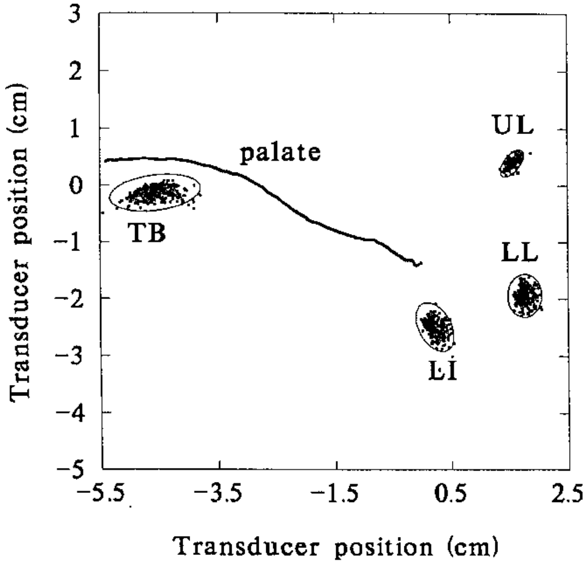




Evidence supporting auditory goals: Motor equivalence in production of /u/



Negative correlations between tongue-body raising and lip protrusion



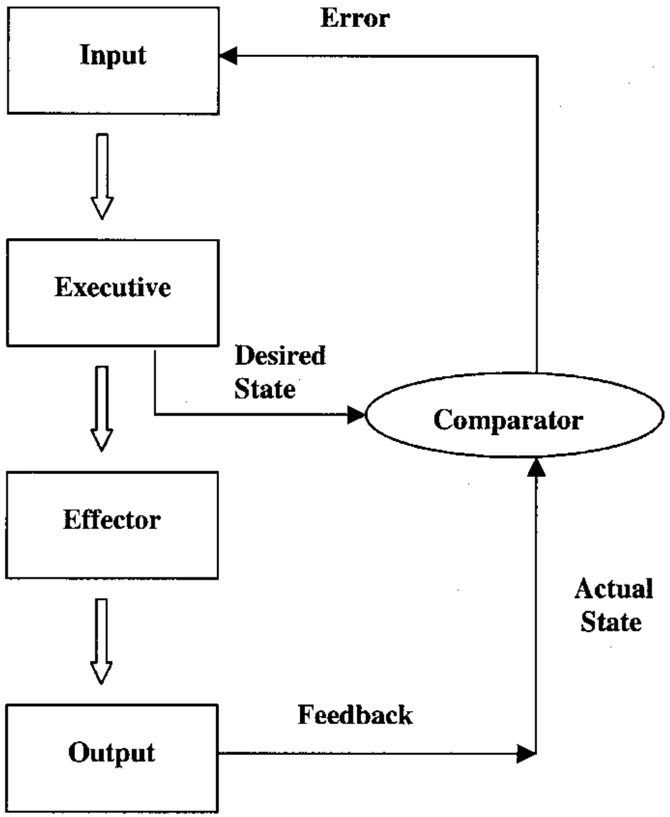
Thus, the goal for the articulatory movements is in acoustic/auditory frame of reference.

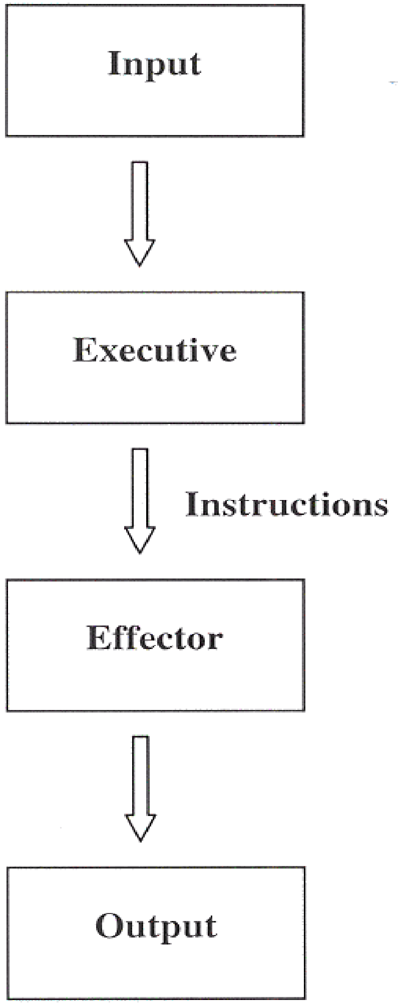
- Perkell, 2003

Feedback vs. feedforward systems

Adapted from: Schmidt, R. A. & Wrisberg, C.A. (2000). Motor Learning and Performance (2nd ed).

Champaign, IL: Human Kinetics.



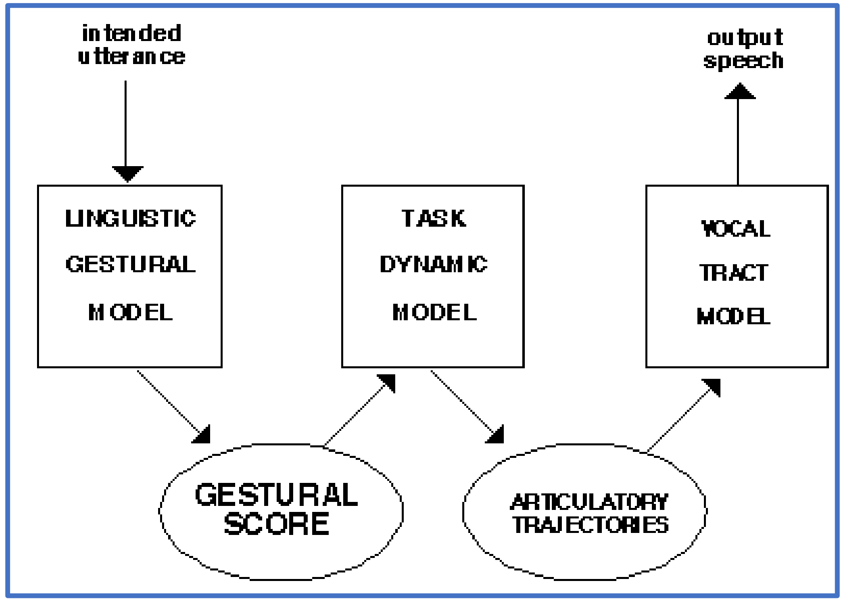


Dynamic Systems Theory (Action theories)

- Motor acts are task-specific
- Motor control works via coordinative structures (synergies)

EXAMPLES:

- Lip closure, opening
- Velar lowering
- Tongue fronting, backing



- Haskins, Gestural model

Dynamic Systems Theory (Action theories) – cont’d

- Gestural theory - assumes phonology is encoded in abstract articulatory gestures
- Can explain magnitude of movements (e.g., rate and stress effects), syllable organization, effects in disordered speech

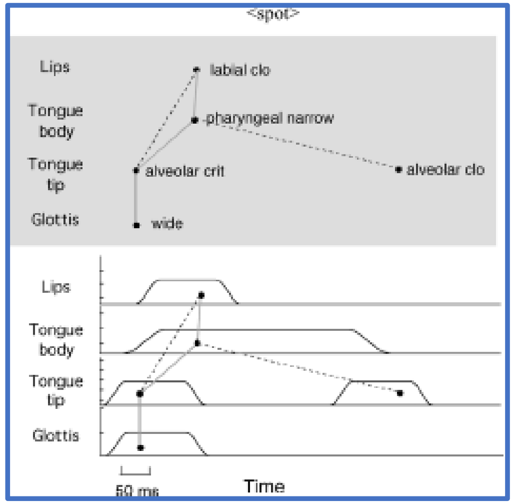


Figure 7.8 The coupling graph for “spot” (top) in which the tongue tip (fricative) gesture and the lip closure gesture are coupled (in-phase) to the tongue body (vowel) gesture, while they are also coupled to one another in the anti-phase mode.

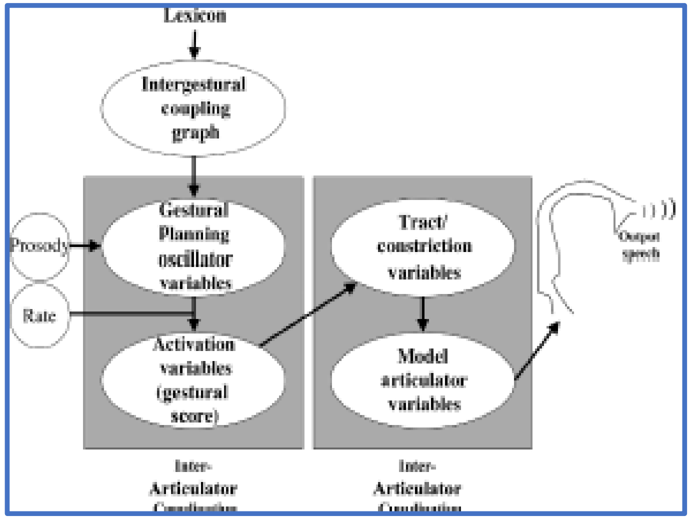
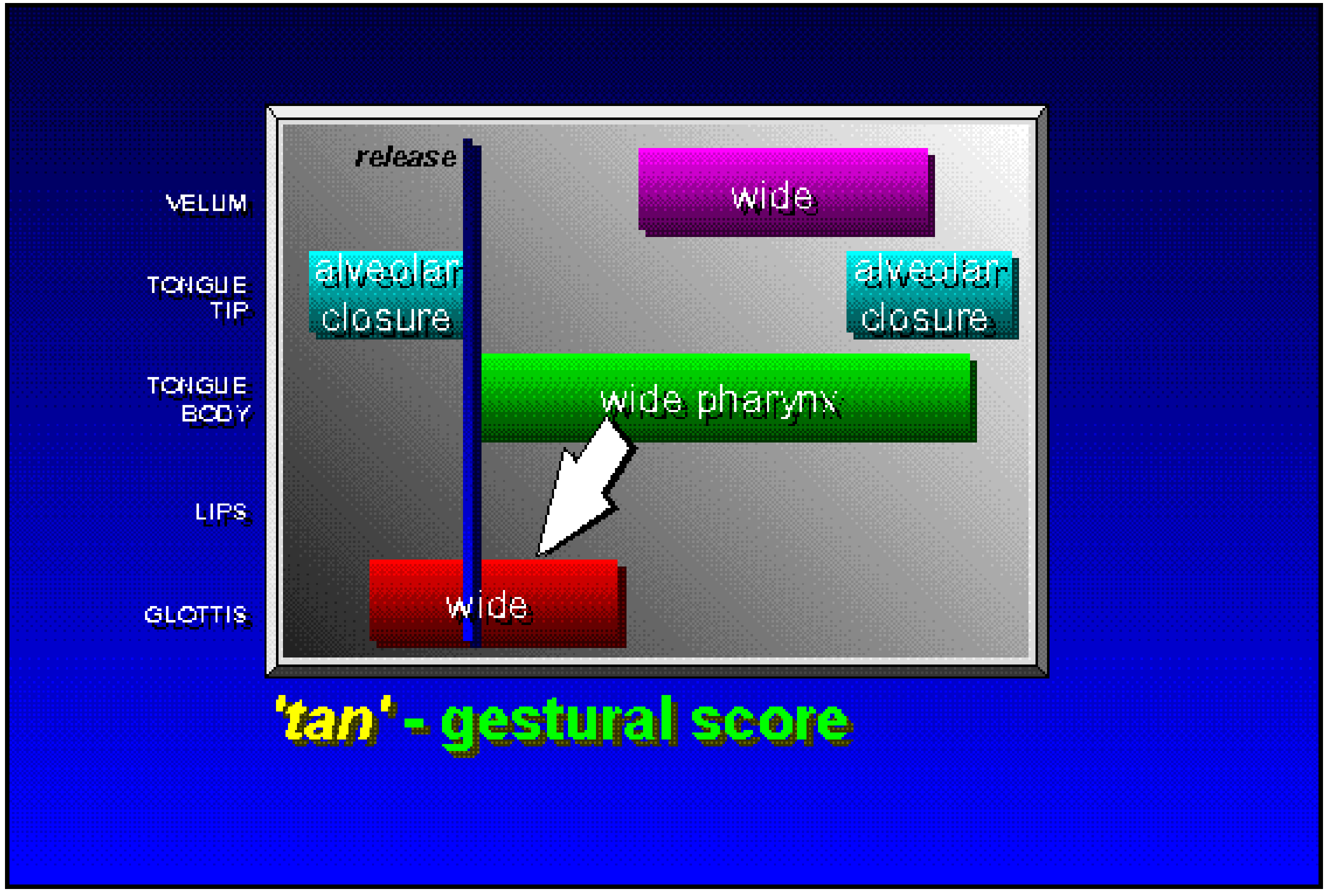
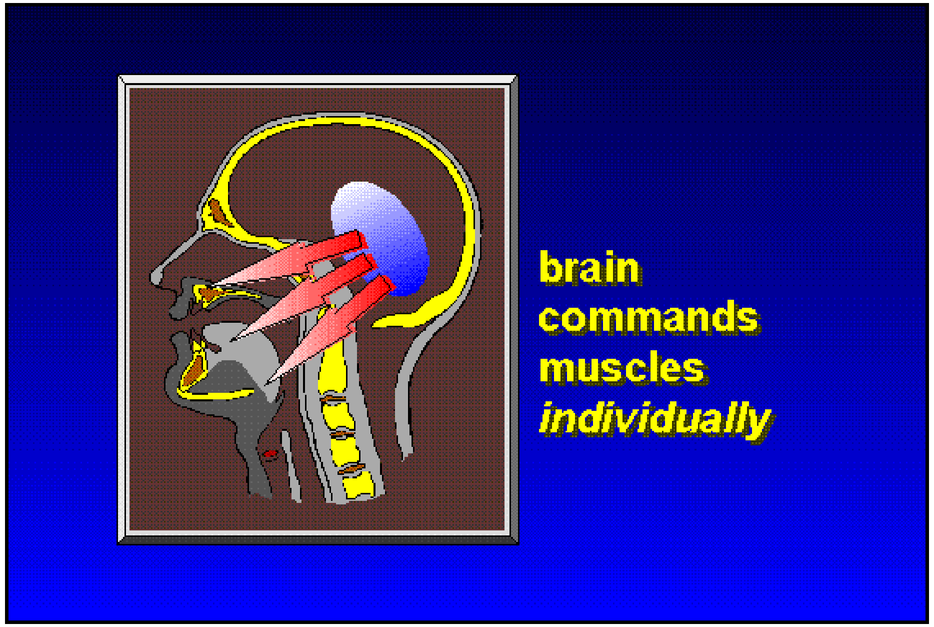

Lexicon
Intergestural coupling graph
Prosody
Rate
Gestural Planning oscillator variables
Activation variables (gestural score)
Tract/ constriction variables
Model articulator variables
Output speech
Inter-Articulator Coordination
Inter-Articulator Coordination

Figure 7.2 The organization of the task-dynamic model of speech production (Saltzman and Munhall, 1989; Browman and Goldstein, 1992; Nam and Saltzman, 2003).

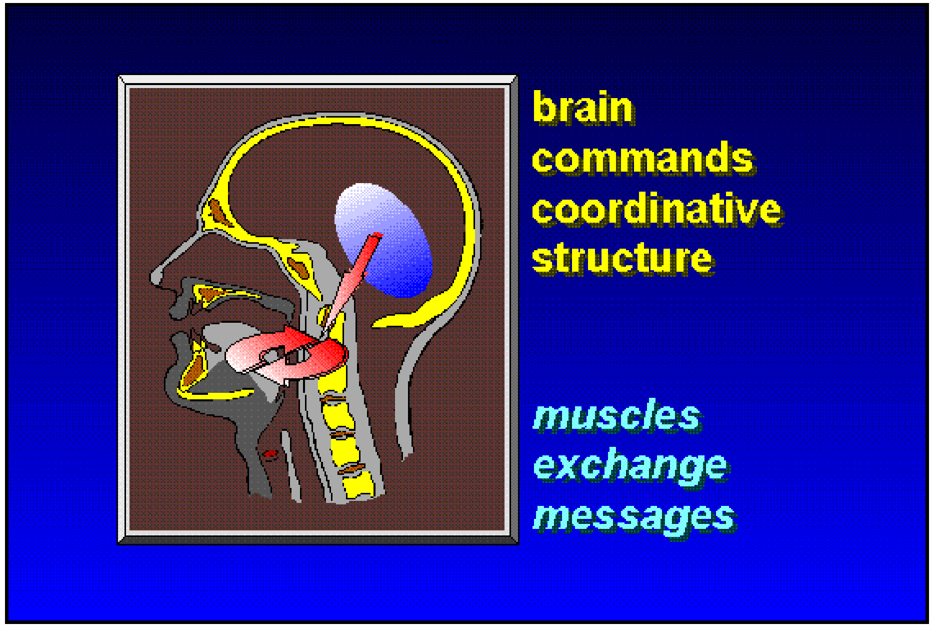
Gesture score - example



Gesture theory - gist



brain
commands
muscles
individually

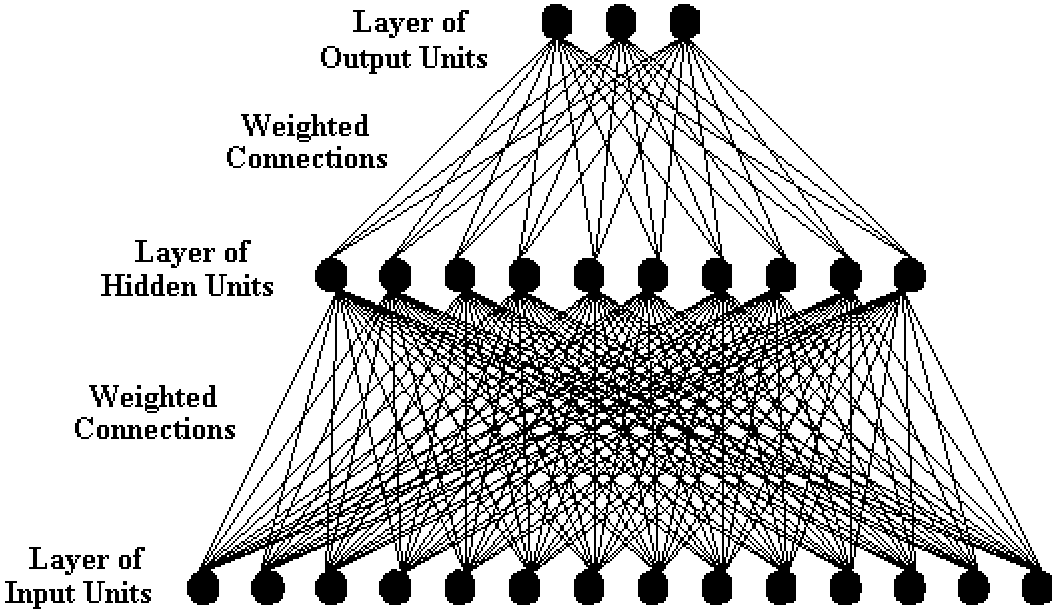


brain
commands
coordinative
structure

muscles
exchange
messages

Old view --------→ New view

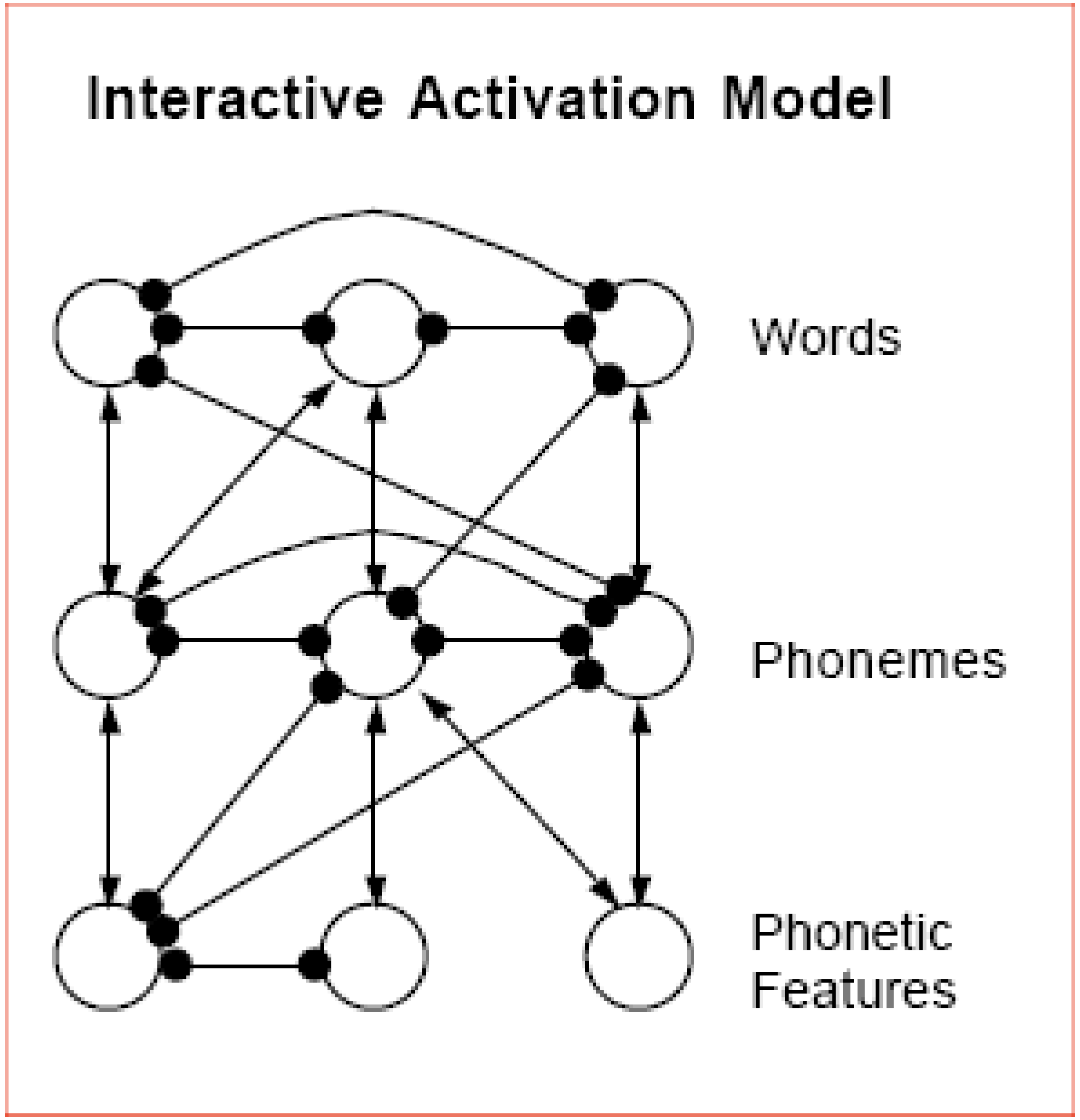
Connectionist Theories (or PDP models)



Parallel Distributed Processing models (PDP)

- Processing = interactions of a large number of simple processing elements called *units*, each sending excitatory and inhibitory signals to other units
- Knowledge → represented in the strength of connections between units in the network
- A concept is not stored, but exist only as long as a particular pattern of activation persists
- What **is** stored is the ability to regenerate that pattern, given an appropriate input cue
- Learning is the development of the right connection strengths

Sample application to speech



The appeal of PDP models

- Seem more closely tied to the physiology of the brain than are other kinds of information-processing models.
- Potentially offer a computationally sufficient and psychologically accurate account of human cognition
- Have altered the way we think about the time-course of processing, the nature of representation, and the mechanisms of learning.

Some quick tutorial links..

1. A very basic introduction

2. A bit more detail

3. Learn from a pro – Geoffrey Hinton (2016) -Neural Networks for Machine Learning

Potential problems of PDP (?)

Too powerful – how to constrain?

Only like the brain in a “toy” sense?

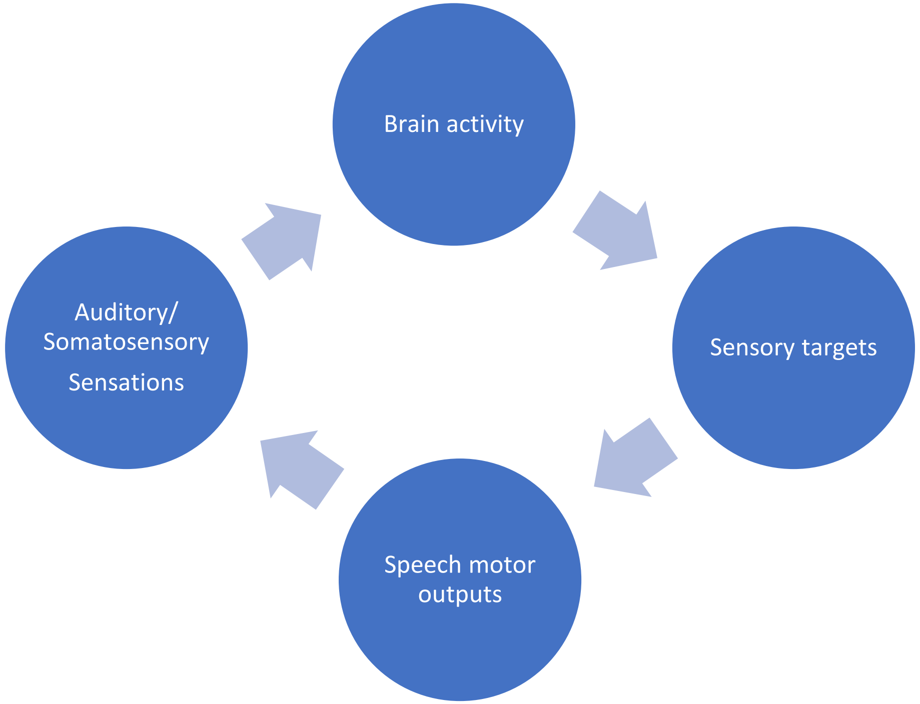
Do these systems really “learn?”

Just because a mathematical model can simulate human behavior, do we really have to believe that humans work that way?

DIVA model

‘Directions Into Velocities of Articulators’

Frank Guenther, BU

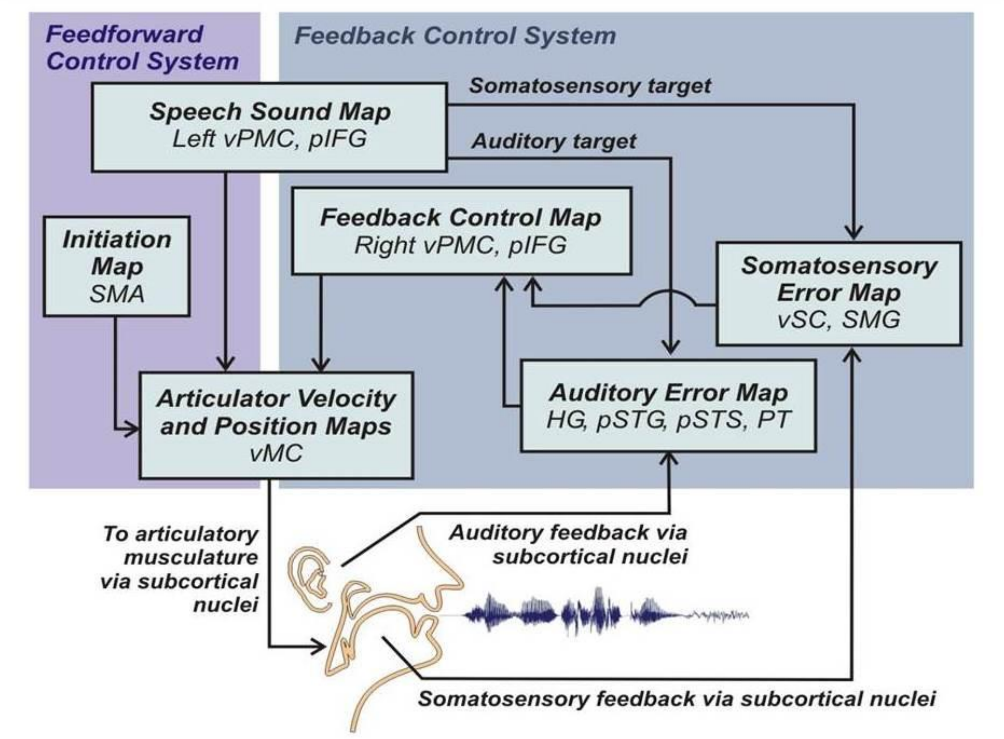


Note:

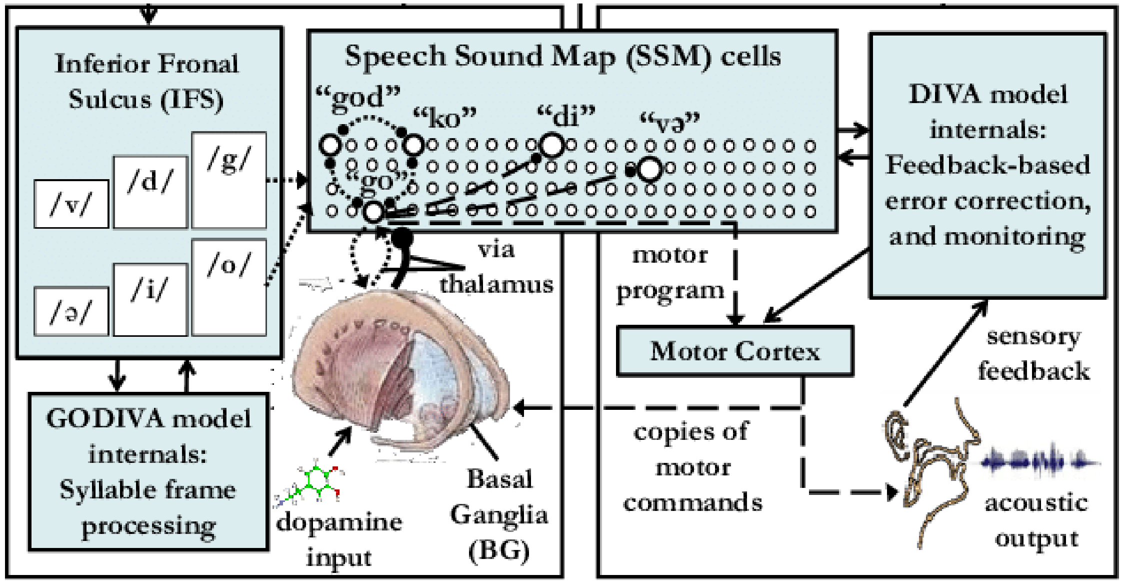
✓ Runs in MATLAB:
We will demo!

Current neurocomputational models

DIVA



GODIVA



Inferior Fronal Sulcus (IFS)
/v/
/d/
/g/
/ə/
/i/
/o/
Speech Sound Map (SSM) cells
"god"
"ko"
"di"
"və"
"go"
via thalamus
motor program
DIVA model internals: Feedback-based error correction, and monitoring
Motor Cortex
sensory feedback
GODIVA model internals: Syllable frame processing
dopamine input
Basal Ganglia (BG)
copies of motor commands
acoustic output

(Civier et al., 2011)

COMD applications

Stuttering, AOS/BA as problems with feed-forward control (feedback is intact)

Problems may be over-reliance on slower feedback routes

Speech perception

SOME HIGHLIGHTS FOR SPEECH LANGUAGE PATHOLOGISTS

A key problem

LACK of invariance

Invariance = perceptual constancy

Listeners effortlessly decode phonemes, but where is this invariant information in the signal?

Other important features of human speech perception

Active process

Depends on clarity of signal

Sound order is important

Signal is fleeting and impermanent (*not* like print)

Listeners can show flexible processing

Speech Perception - Key Concepts

Bottom-up vs. Top-down

Active vs. Passive

- (Very similar to “controlled” vs. “automatic”)

Autonomous vs. Interactive

Bottom up processes

Information comes in from periphery to central processing systems

⬆

Example:

Ear → *auditory nerve* → *primary auditory cortex* → *Wernicke’s area* → *greater peri-sylvian cortex*

Top-down Processes ⬇

Processes whereby pre-existing knowledge sources are brought to bear on the decision as to what speech sounds are being heard.

Examples:

VOT boundaries shift based on whether a sound is a word in the language:

- ***"boat"*** and not "*poat*"
- ***"cope"*** and not "*gope*"

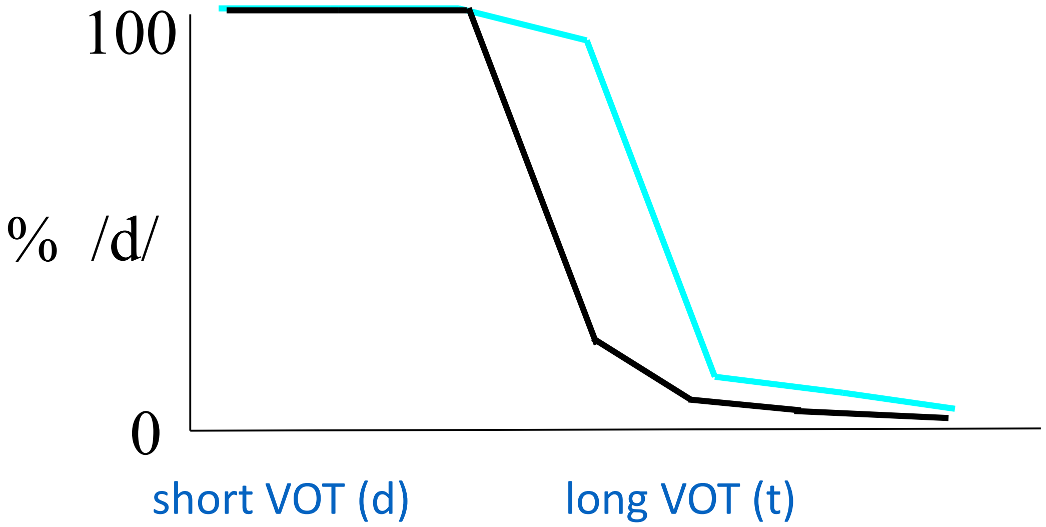
(see next slide →)

Top-down processing in speech perception

- VOT identification experiment
- Word at one end, non-word at the other
- Perception is more forgiving when the sound means something!

nonword-word: dask-**task**

word-nonword: **dash**-tash



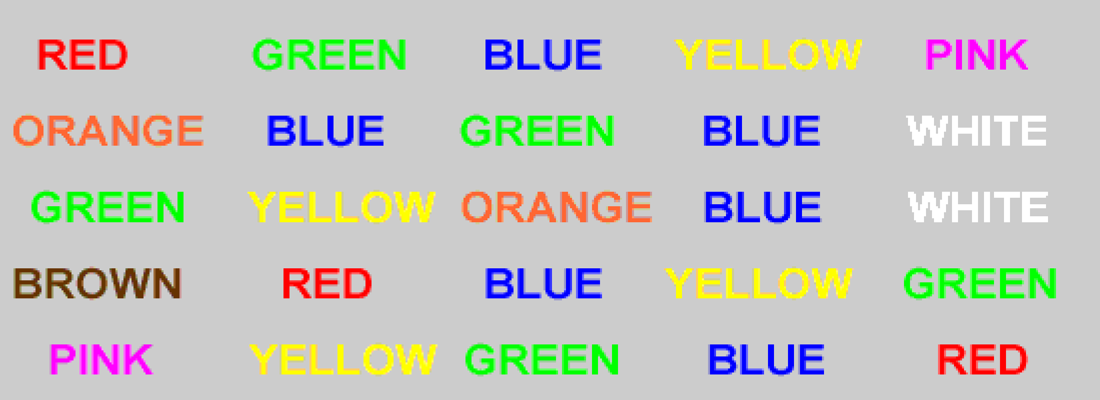
- Ganong (1980)

Active vs. Passive Processing

Active (or controlled) requires processing resources (e.g. attention), can be manipulated

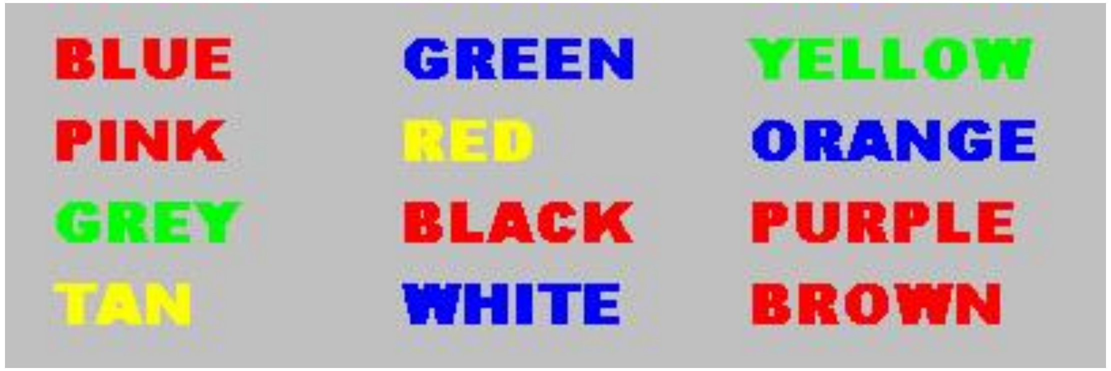
Passive (or automatic) is relatively effortless and not affected by external conditions

Example of automatic vs. controlled: Stroop effect



RED	GREEN	BLUE	YELLOW	PINK
ORANGE	BLUE	GREEN	BLUE	WHITE
GREEN	YELLOW	ORANGE	BLUE	WHITE
BROWN	RED	BLUE	YELLOW	GREEN
PINK	YELLOW	GREEN	BLUE	RED

Fast, accurate



BLUE	GREEN	YELLOW
PINK	RED	ORANGE
GREY	BLACK	PURPLE
TAN	WHITE	BROWN

Slow, inaccurate

Autonomous vs. Interactive

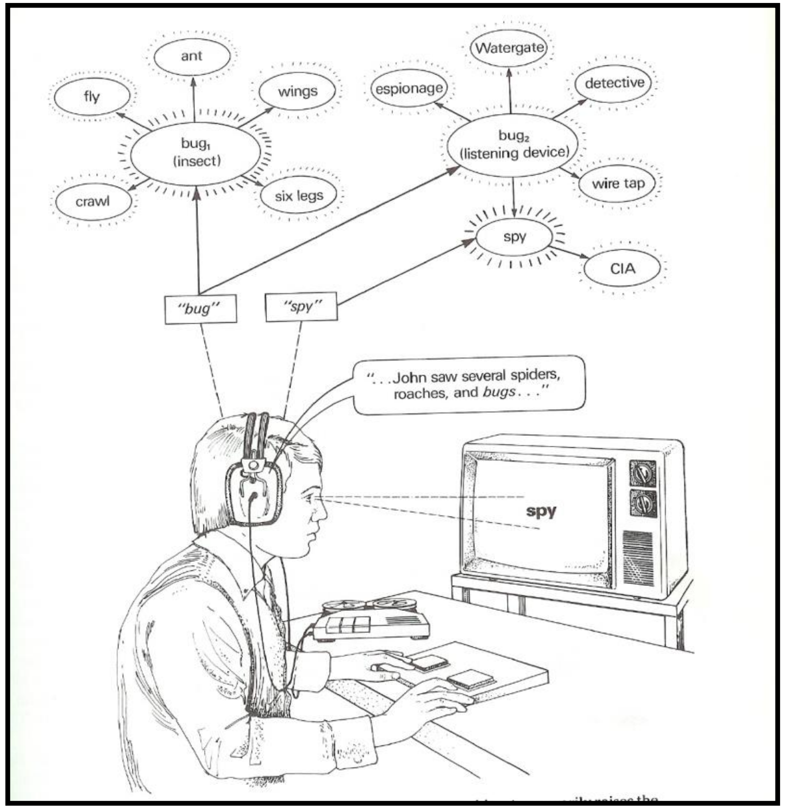
Autonomous – closed system of decision making (capsulated, modular)

Interactive – decision-making process relies on various sources of information outside the perceptual processor

Classic experiment - illustrates autonomous vs. interactive properties (..for *lexical* access)

“*Lexical Access During Sentence Comprehension: (re) Consideration of Context Effects*” -- D. Swinney (1979)

- An online measure – “Cross-modal priming”
- Subjects hear sentences containing ambiguous words (e.g. BUG) while seated in front of a computer screen. At the same moment the ambiguous word is uttered, a simultaneous string of letters, either a word or a non-word, is flashed on the computer screen.
- These words reflect one or another meaning of the ambiguous word (e.g. ANT, SPY) or are unrelated controls (e.g. SEW).
- Subjects respond as quickly as possible – by hitting a button - once the probes were processed.
- The idea is that multiple meanings are activated at the moment an ambiguity is encountered -- priming related concepts.



"*Lexical Access During Sentence Comprehension: (re) Consideration of Context Effects*" -- D. Swinney (1979)

Motor theory of speech perception

“Speech is understood in terms of how it is produced”

- Speech relies on an auditory code
- The auditory signal reflects complex encoding
- Nevertheless, the claimed invariant is articulatory
- That is, motoric gestures must be recovered from the acoustic signal.
- Examples include the “locus” for stop consonant place of articulation.

Motor theory

EVIDENCE FOR:

Duplex perception

Mirror neurons?

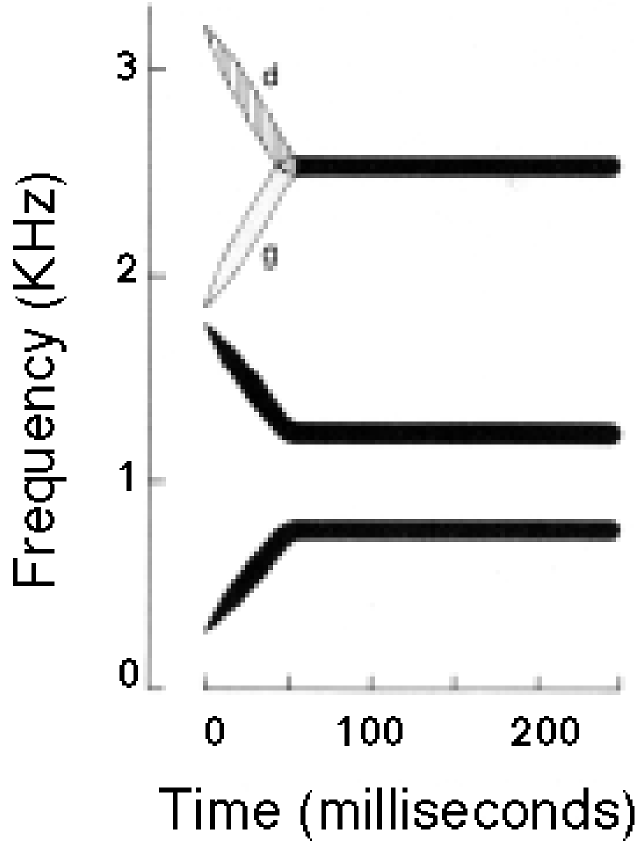
EVIDENCE AGAINST:

Comprehension precedes production (e.g. in development)

Parsimony

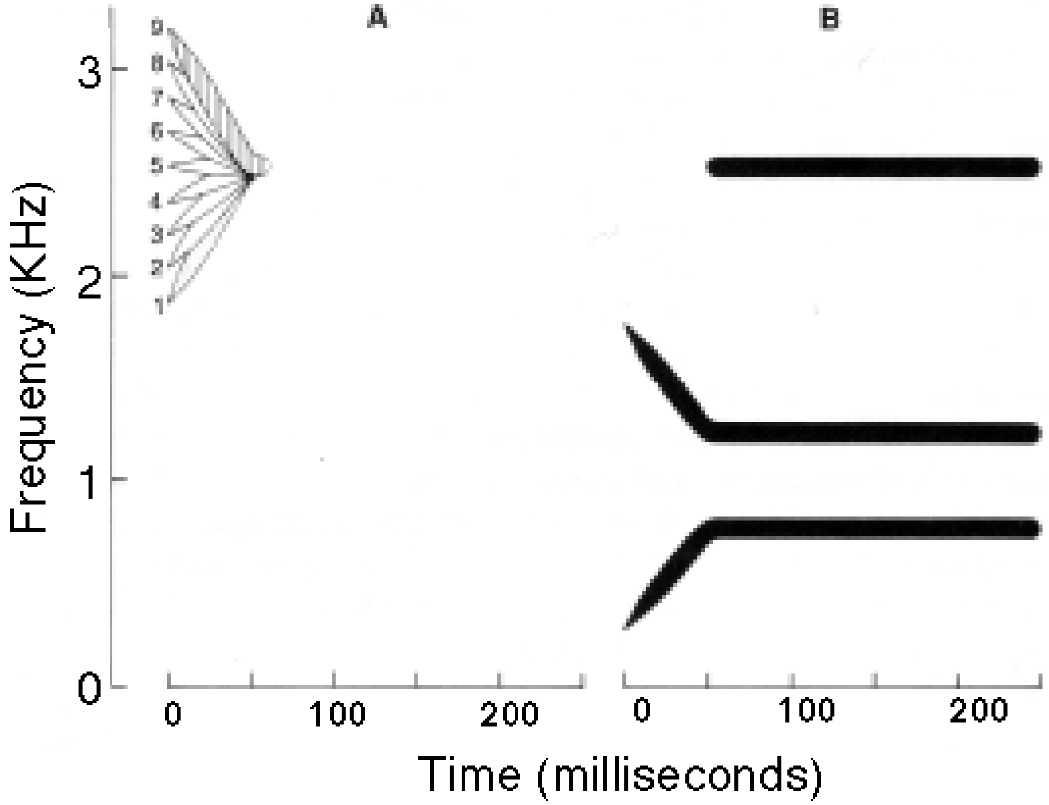
Duplex perception - introduction

Experiments make use of stimuli in which direction of F3 transitions distinguish [da] from [ga]. Without this transition, the rest of the stimulus pattern is ambiguous between [da] and [ga].



Duplex perception - methodology

The critical formant transition (A) is presented to one ear, and everything else (the ambiguous "base", B) is presented to the other.



Duplex perception - results

WHAT IS HEARD:

- Listeners hear a syllable in the ear that gets the base (B). Its identification is determined by which of the nine F3 transitions are presented to the other ear (A).
- Listeners also hear a non-speech "chirp" in the ear that gets the isolated transition (A).
- Thus, perception is “duplex”

Duplex perception - implications

Same stimulus is simultaneously part of two distinct types of percepts

Thus, percepts are produced by separate mechanisms, or modules, that are both sensitive to the same range of stimuli.

The discrimination functions for the isolated "chirp" and the speech percept are quite different.

The speech percept exhibits categorical perception, the chirp percept exhibits continuous perception.

Duplex perception

Concerns…..

“Why does this necessarily support the motor theory?”

Other theories of speech perception

- Acoustic invariance theory
- Direct realism
- TRACE model
- Logogen theory
- Cohort theory
- Category decision models: Prototypes, fuzzy logical models, native language magnet theory

For more information on these other speech perception models, see:

http://www.utdallas.edu/~wkatz/courses/More_info_speech_perc_models.pptx

NOTE: (Optional and for fun! Not needed for final exam)

McGurk Effect

MacDonald & McGurk (1976)

- Visual modality may complement or even override auditory input
- Effects are complex -- the basis for these effects remains controversial
- See e.g. http://www.youtube.com/watch?v=T4fUi0eG1X4
- OR http://www.youtube.com/watch?v=aFPtc8BVdJk&feature=related

Clinical import:

- Lip reading and/or facial cues in deaf language
- Added difficulty of non-visual modes of communication (e.g., the telephone) for communication disordered patients

Life course issues

- Young infants may be equipped with a “universal phonetic analyzer”
- By ~ 1 year old, infants have already tuned in to their native language(s)
- Young children’s perception appears remarkably adult-like at very young ages; chief difference seems to be increased variability. However, more research needed in these areas.
- Morphemes with “low phonetic substance” (e.g., *-ed*, *-s*) may be particularly affected in children with specific language impairment (SLI)
- Speech perception seems to be well maintained in normal aging