Phonological learning: auditory, visual, computational and animal perspectives

William Idsardi, University of Maryland

GLOW 2019

Thanks to:

Workshop organizers: Julian Lysvik and Andrew Nevins

All the speakers

The audience

GLOW

University of Oslo

Oslo and Norway

Outline

Overall goals

The EFP framework

Temporal underspecification

From syllables to segments in spoken language development

Animal models for temporal order

Symmetry in sign language: Speculations

Overall goals

A definition for phonology (what's being learned?)

A general framework for MAP systems

Explicit representation of precedence

Logic, model theory, low computational complexity

Comparisons across modules, systems, species

Modal, amodal, multi-modal

Anderson 1993: “proposed commonalities”

“Linguistic expression and its relation to modality”

“A point that seems often to constitute an important subtext, although seldom presented as an explicit argument, is the observation that both signed and spoken language expression systems can be analyzed in terms of the formalism of auto-segmental (and metrical) representations (cf. Goldsmith, 1990). I submit, how ever, that this is a fact which by itself is of singularly little interest, at least for the conclusion that the two kinds of system are substantively similar. The reason for this is that when we examine the nature of autosegmental (and metrical) repre sentations, they turn out to be so general that it is hard to imagine any sort of skilled movement that would not fall into this class.”

EFPS

Events, features, precedence, spatial

The EFP framework

Spoken language: <E,F,P> Events, Features and Precedence

Events are abstract points in time, **Features** are monadic properties of events (like [round]), and **Precedence** is a dyadic temporal relation between events (“maybe next”, after, open bigrams)

Sign language: <E,F,P,S> Events, Features, Precedence and Space

Events are abstract points in spacetime, **Features** are monadic properties of events (like [5]), **Precedence** is a dyadic temporal relation between events, and S is a set of dyadic **Spatial** relations between events (such as symmetries)

Features in a Memory-Action-Perception loop

Features are the memory “glue” between production and perception

Jakobson, Fant & Halle (but not Morris’s later views)

Formants drop (STRFs) ⇔ [round] ⇔ orbicularis oris innervation

Right angle in hand ⇔ [pointer] ⇔ extensor indicis innervation

- Infant recognition of biological motion in point-light displays

The production and perception primitives are not specific to language (pre-exist).

The **glue** makes them “special”. (MAP generally?)

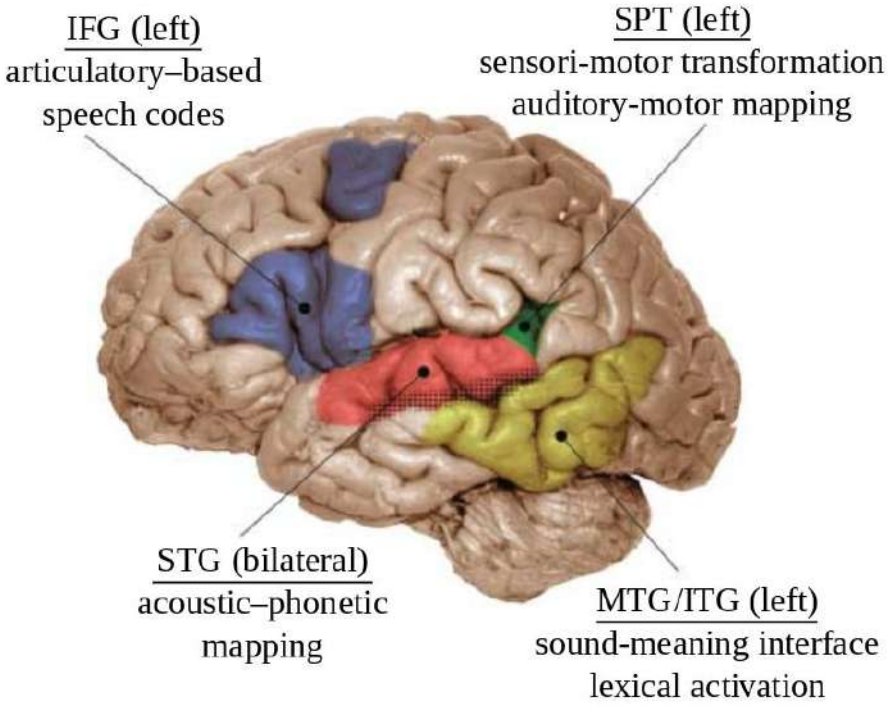
Proto-maps (Science last week)

The mammalian brain's somatosensory cortex is a topographic map of the body's sensory experience. In mice, cortical barrels reflect whisker input. We asked whether these cortical structures require sensory input to develop or are driven by intrinsic activity. Indeed, thalamocortical columns, connecting thalamus to cortex, emerge before sensory input and concur with calcium waves in the embryonic thalamus. ... Thus, a self-organized **protomap** in the embryonic thalamus drives functional assembly of murine thalamocortical sensory circuits.

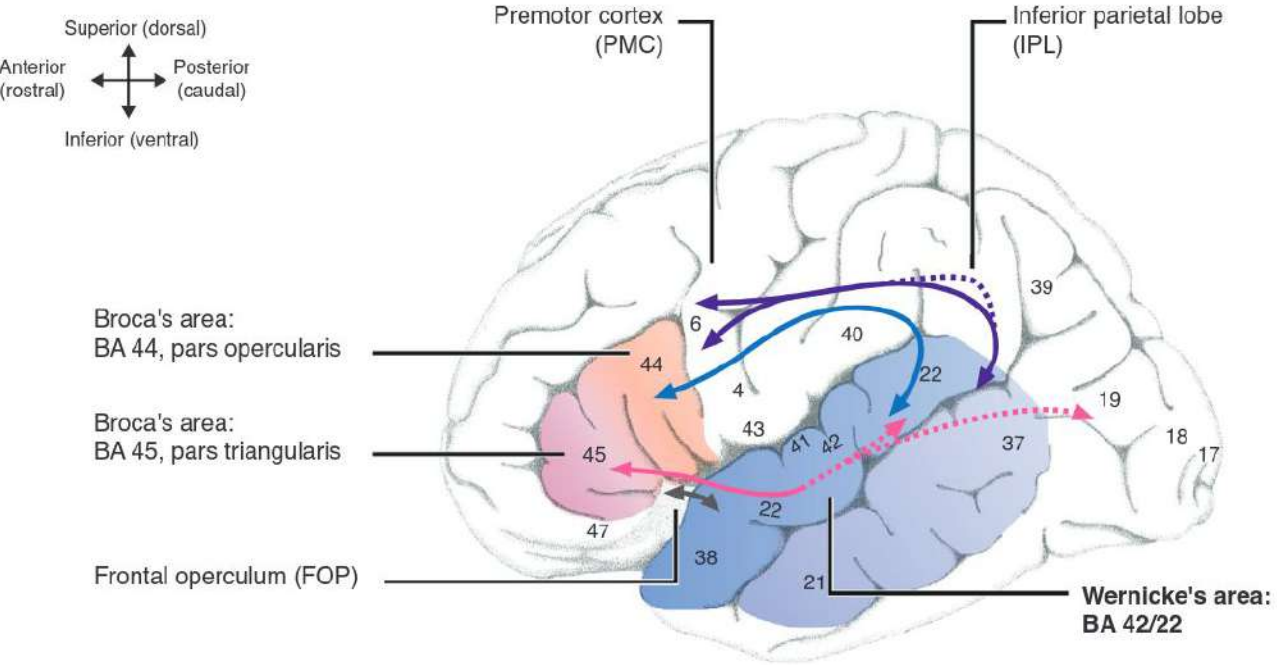
Dual-streams (Hickok & Poeppel)

Events, features and precedence must be mapped between systems

Coordinate transforms



Friederici 2018



Dorsal fiber tracts

- PMC to pSTG
- BA 44 to pSTG

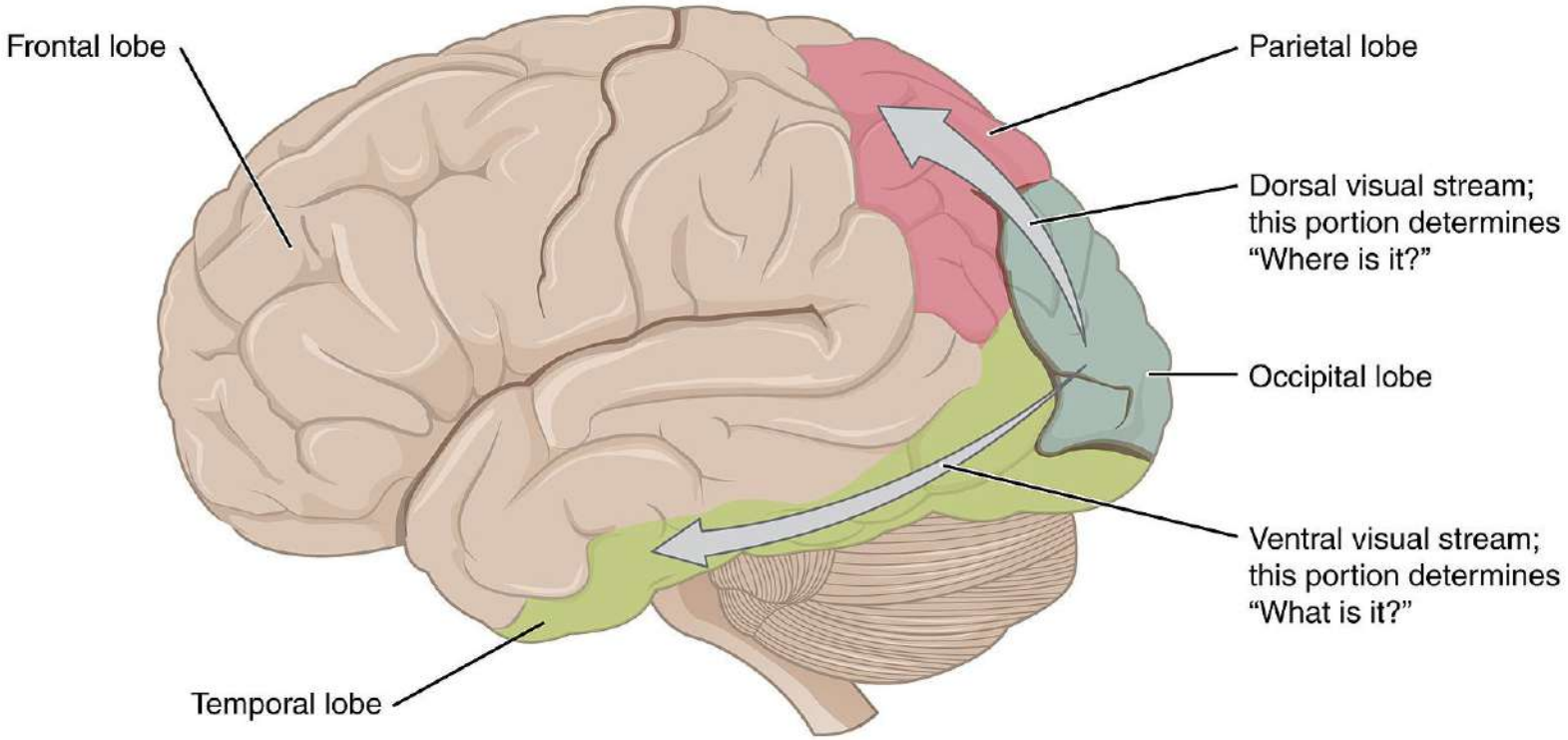
Ventral fiber tracts

- BA 45/47 to STG/MTG
- FOP to aSTG

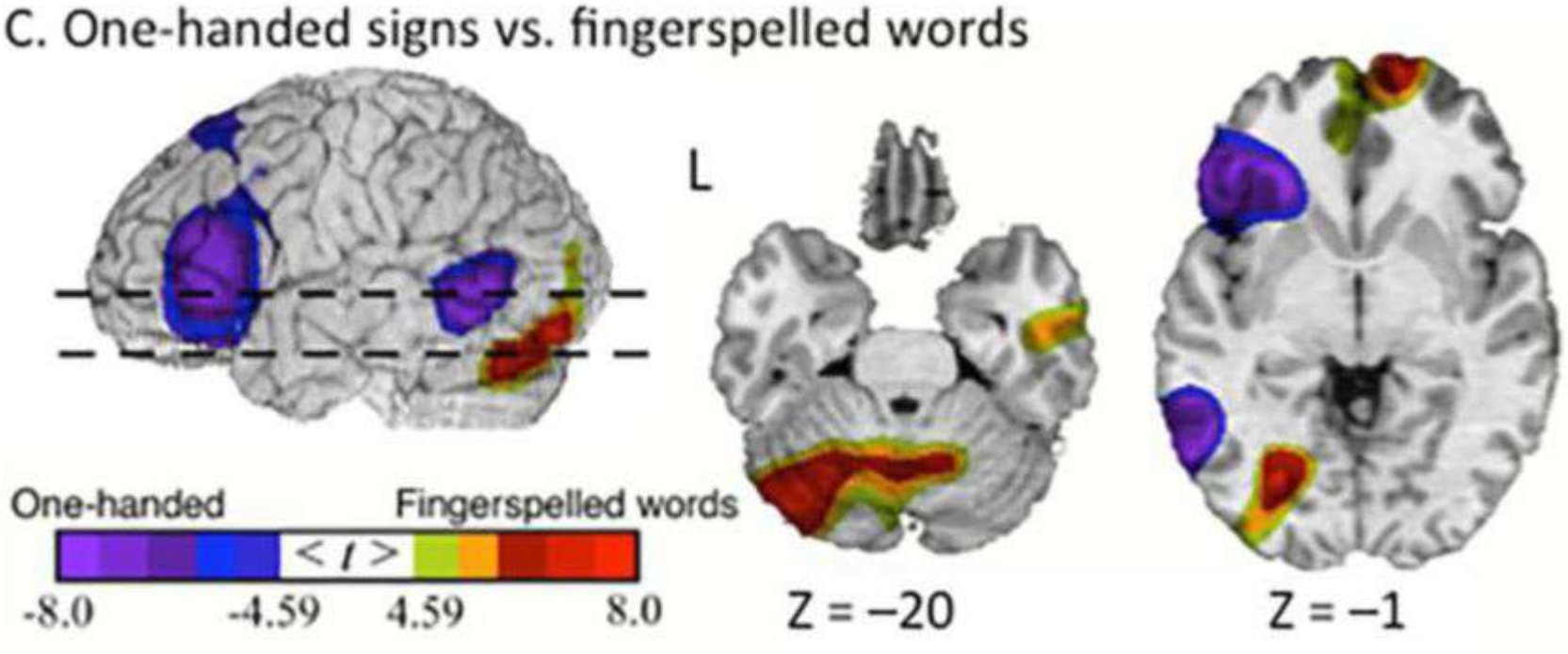
- Inferior frontal gyrus (IFG)
- Superior temporal gyrus (STG)
- Middle temporal gyrus (MTG)

Current Opinion in Behavioral Sciences

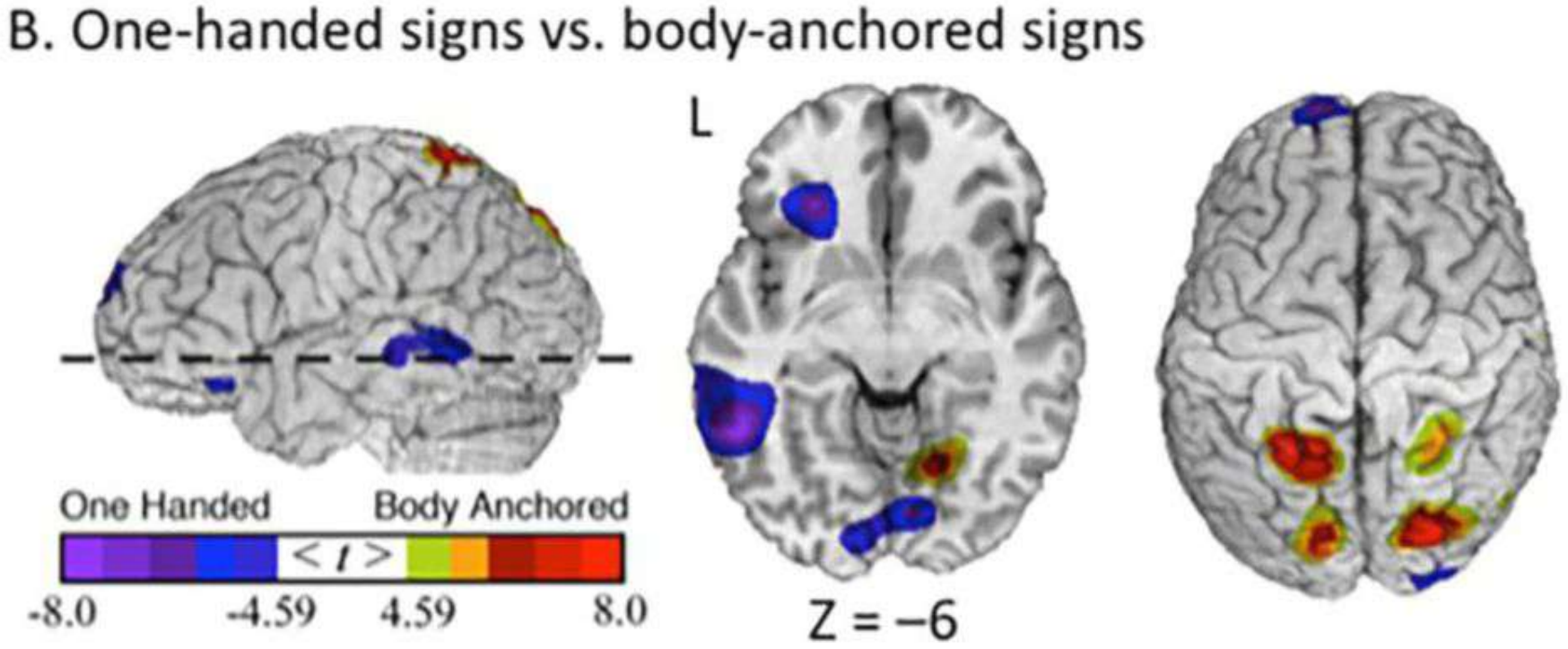
Dual streams in vision (Wikipedia)



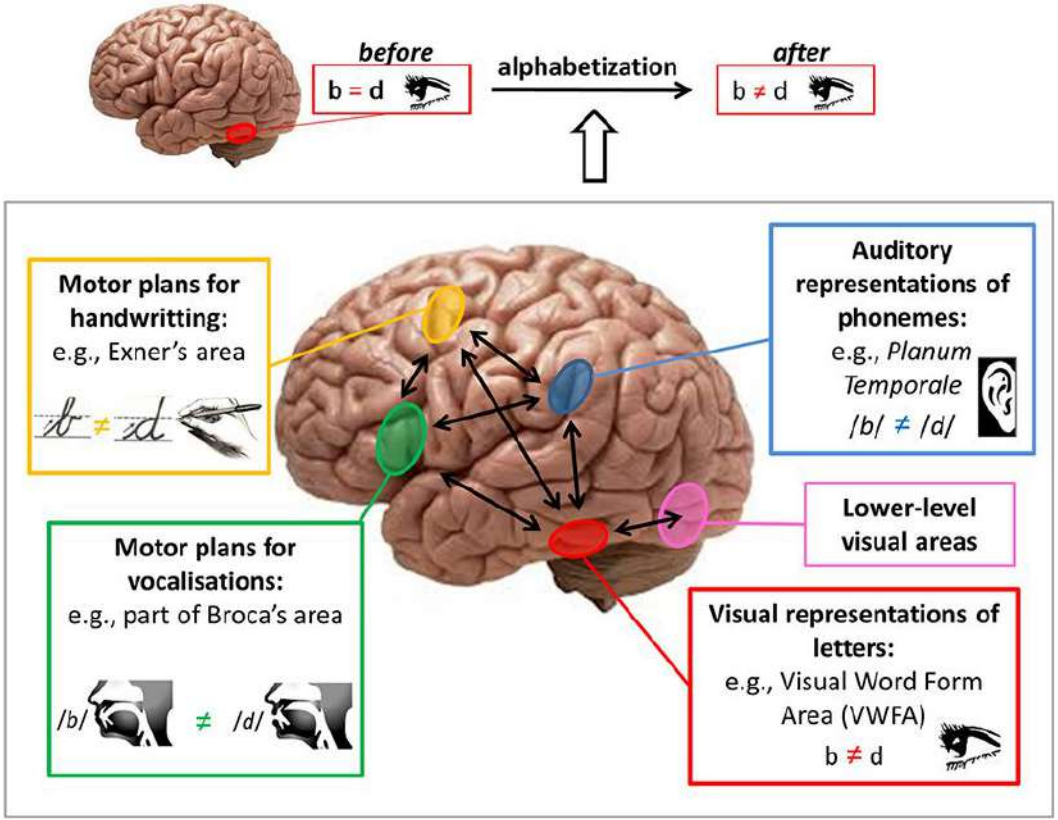
Sign vs finger-spelling (Emmorey et al 2016)



Different sign types (Emmorey et al 2016)



Quick overall picture from reading (Wikipedia)



Challenges for sign EFPS

What are the degrees of freedom in the control of the arm and hand?

How many frames of reference? (finger, hand, wrist, elbow, shoulder, body)

What visual feature detectors are there? Biological motion? Object-centered?

How do they match up?

(This looks really difficult)

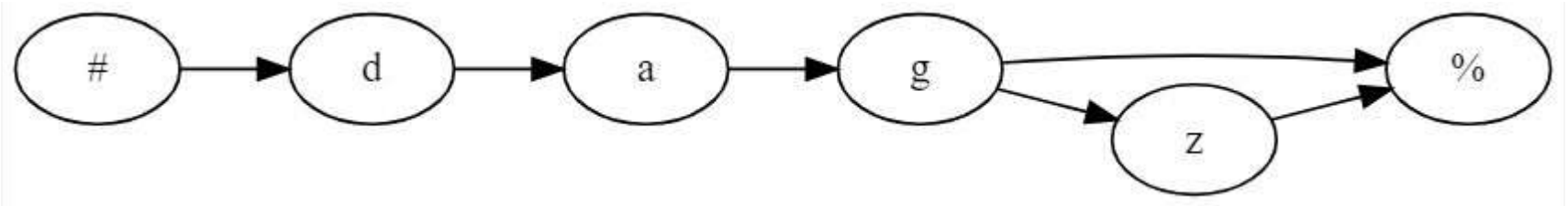
Some sign features (Keane et al 2017)

group	feature	-C-	-A-
primary selected fingers	members	index, middle, ring, pinky, thumb	index, middle, ring, pinky
	base (MCP) joint	extended	flexed
	non-base (PIP, DIP) joints	mid	flexed
	abduction	adducted	adducted
secondary selected fingers	members	*none*	thumb
	base (MCP) joint	*n/a*	mid
	non-base (PIP, DIP) joints	*n/a*	extended
thumb	opposition	opposed	unopposed
non-selected fingers	members	*none*	*none*
	joints	*n/a*	*n/a*
wrist	orientation	FS-default	FS-default

Precedence graphs

Phonological representations are <E,F,P> or <E,F,P,S> graphs.

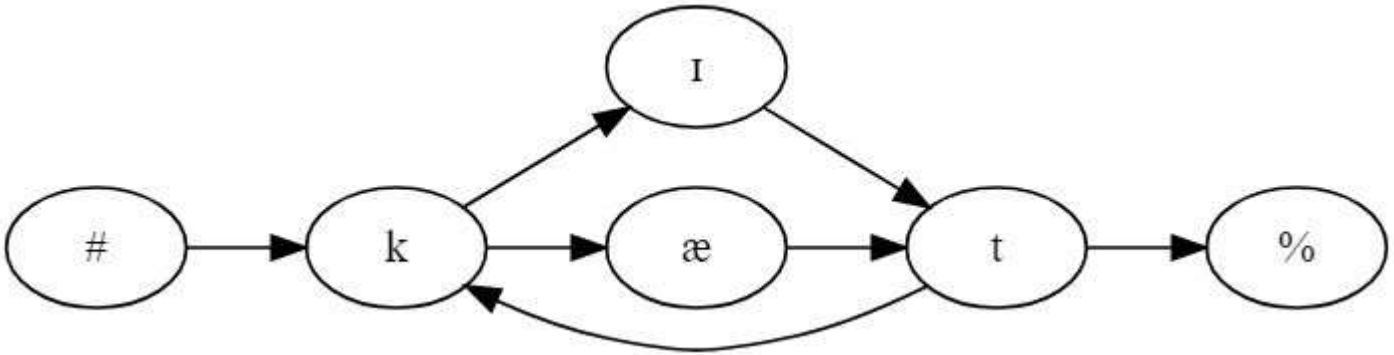
Following Raimy (2000) and Papillon (2018) very few additional stipulations are made. Therefore the graphs are not guaranteed to provide a total order over events.



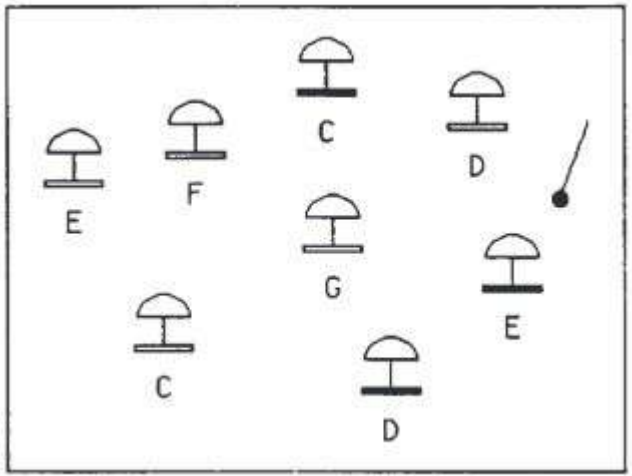
Reduplication

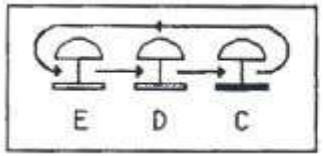
Reduplication is captured with a loop (cycle) in the graph.

kit-kat:



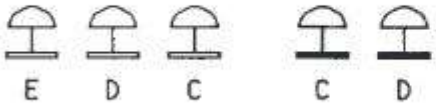
Hot Cross Buns (Bamberger 1991)





b. # → E → D → C → %

a.

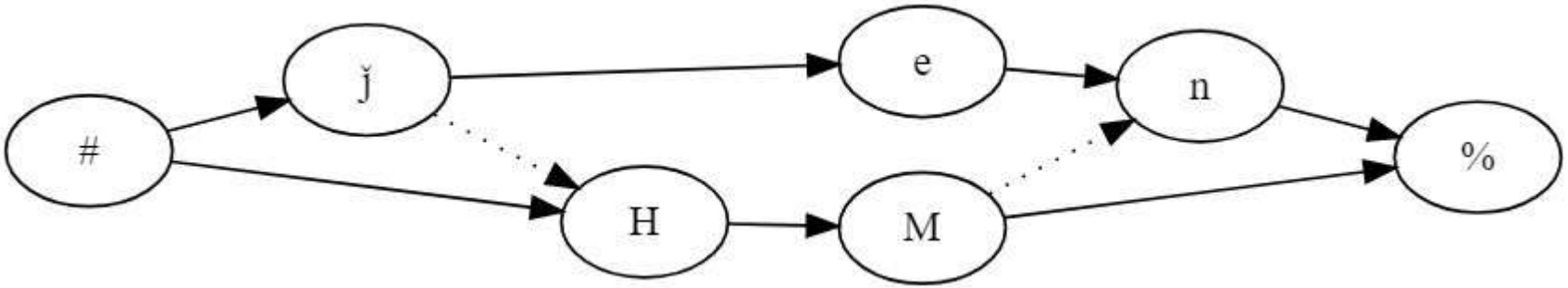


b. # → E → D → C → C → D → %

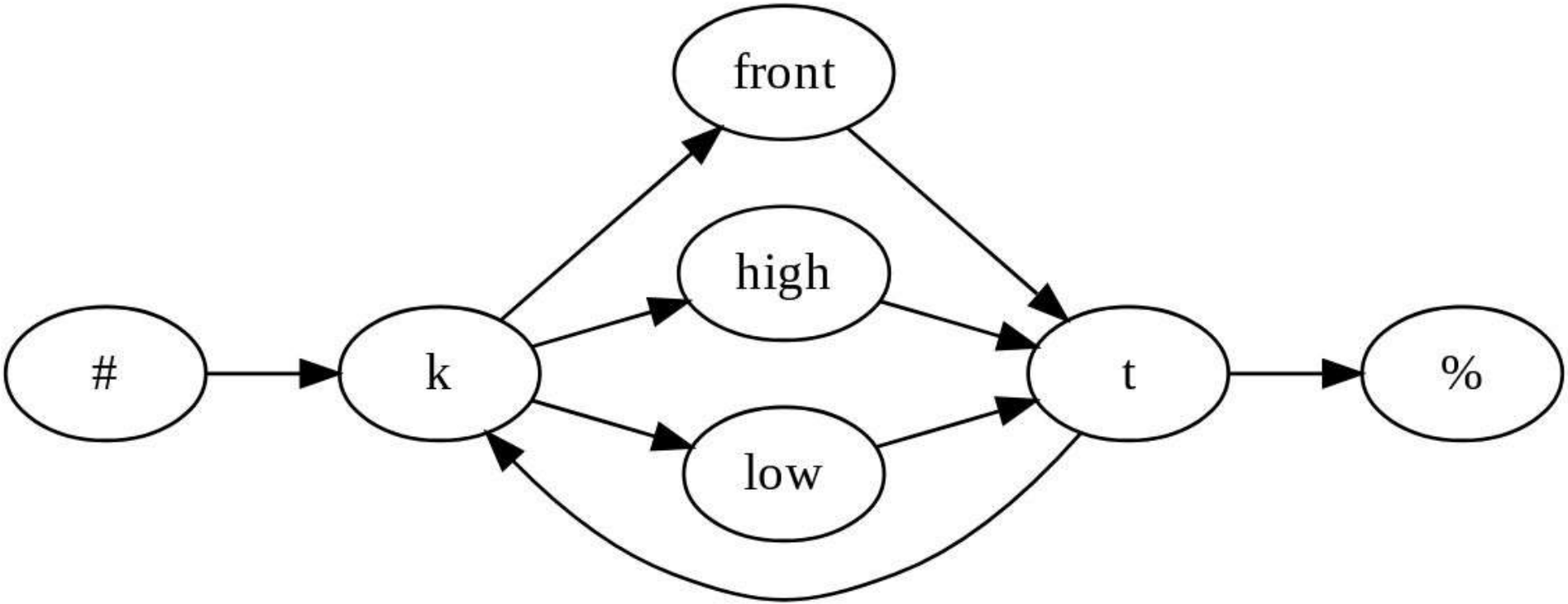
4 4

Parallel paths

If neither a<b nor b<a, then (defn) a||b (a is parallel to b). If a and b are not incompatible events, then they can occur simultaneously (overlapped in time) but they need not be so realized. In perception this is a way to capture Bregman's auditory streams. This is the formal mechanism for notions like tier segregation.



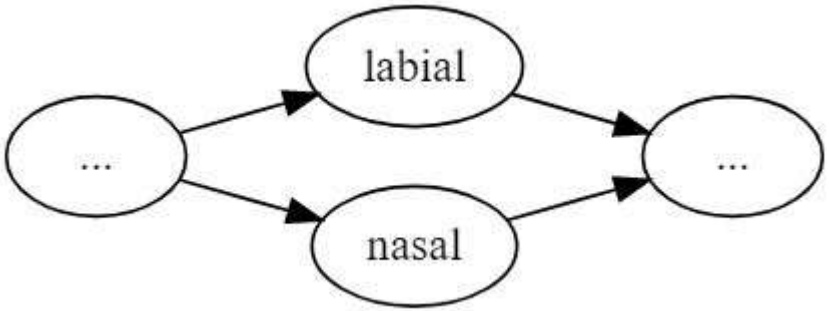
Kit-kat again

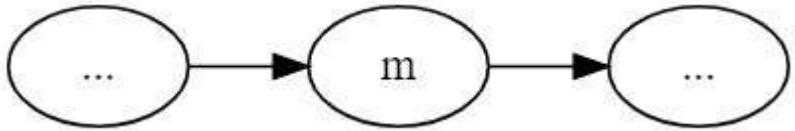


Complex events

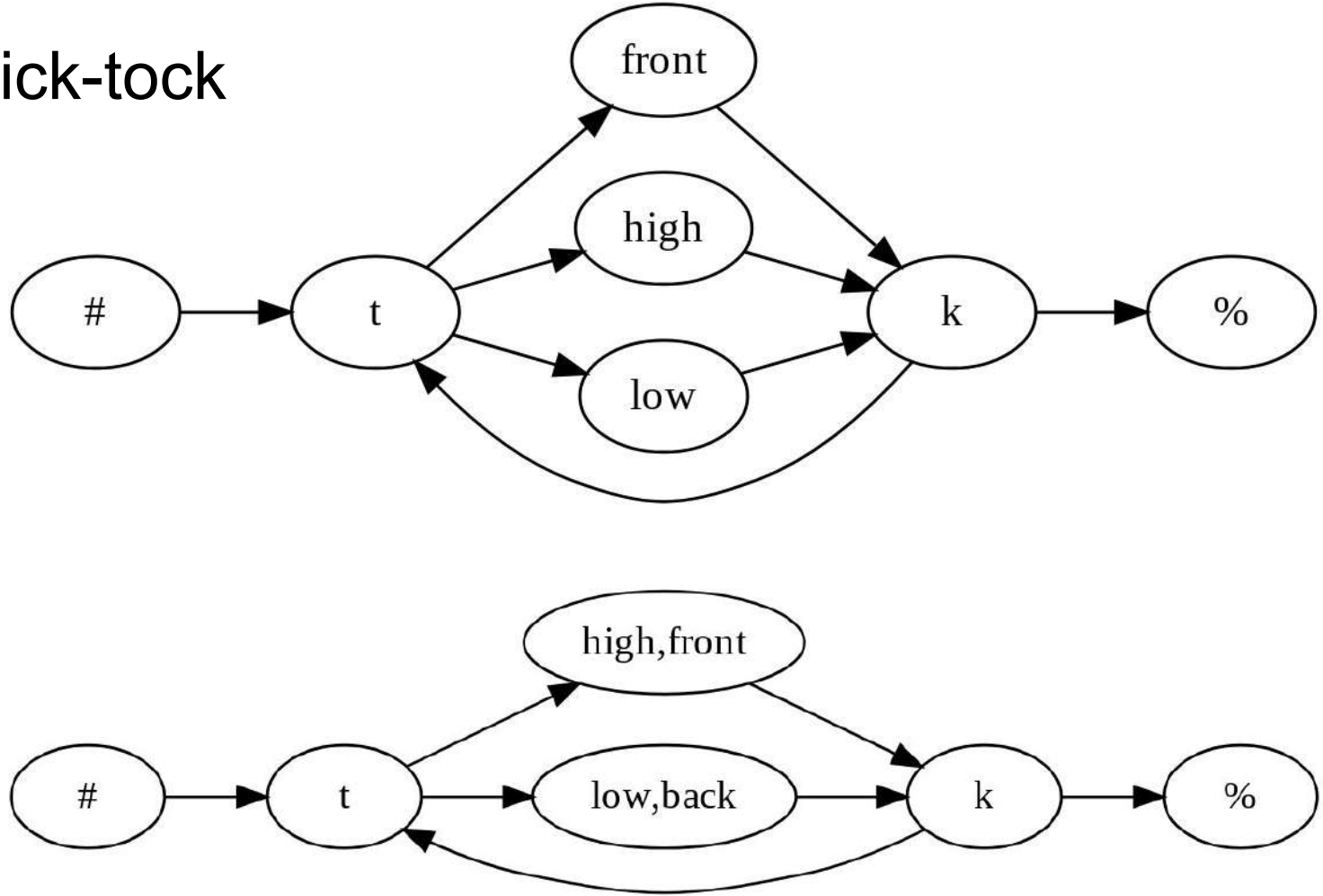
Events can have multiple properties ([F,G]e) this will be one method of coordinating features in time.

The origin of segments is feature binding labial || nasal => m





Tic-tac vs tick-tock



Special events

Some special events: # (beginning) and % (end).

No event before #, no event after %:

For all e, not e < #. For all e, not % < e. (< is precedes)

No features on #, %:

For all F, not [F]#. For all F, not [F]%. (These are MSO.)

Implication: there can be featureless (empty) events (~ junctures)

Why EFP?

Simplicity and model-theoretic interpretability. (Similarity to Pietroski 2018.)

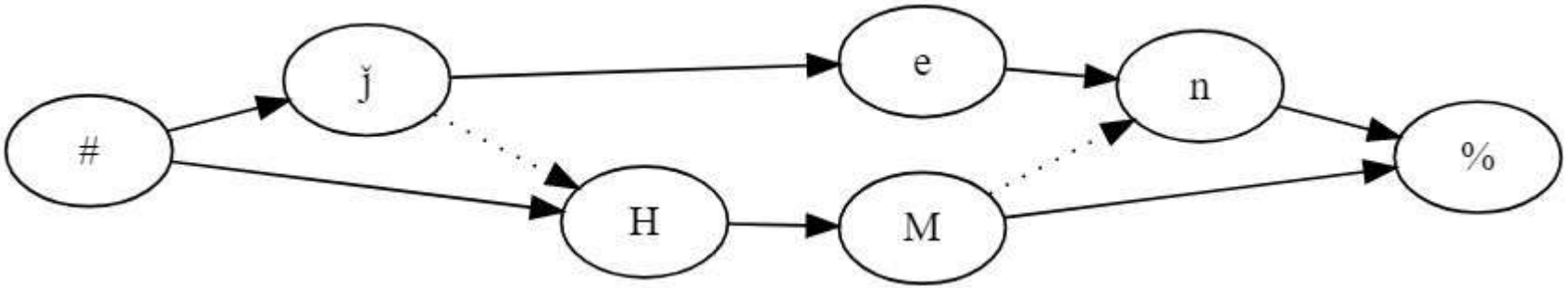
Allows for underspecification in time (and spacetime). This allows for representations to be “built up” derivationally. It allows some temporal decisions to be postponed (or clarified later).

Clarifies the developmental trajectory from “holistic” representations to phonemic ones.

Clarifies the relationship between phonology and “phonology adjacent” systems such as music, birdsong, etc.

Underspecification and elaboration

If (a<b<c)||(x<y<z) this can be further elaborated through subsequent information. This will allow us to gradually transform a representation underspecified in time into one that is more temporally specified.



Underspecification and elaboration

The gradual transformation can be in a particular derivation (production or perception), or can be developmental (infant language learning)

Similar to D-theory (Marcus, Hindle & Fleck 1983), no retraction of facts.

Speech development

From syllables and features to segments

Order perception

Auditory temporal order perception is not that good (Warren 2008).

For pairs, as low as 20-30ms in adults, > 100ms in 5 year olds.

For 4+ sequences > 200ms in adults (~ syllable).

Thus, speech perception is special.

In general, even harder to determine the order across streams (Bregman)

Syllables and features early, no segments

Young infants (< 4 months) show a puzzling set of abilities and lack of abilities

Can detect feature changes (Eimas)

Can recognize syllables but not segments (Mehler, Jusczyk)

Dual time scale (Poeppel) nesting features (gamma) within syllables (theta)

Featural precedence is hard, develops over time

Segments develop through “features coordinated in time” (Kazanina et al 2017)

Multisensory integration

AV speech

In adults: McGurk effect

In young infants:

Kuhl & Meltzoff

Patterson & Werker

(If phonology is “amodal”, is it multi-sensory?)

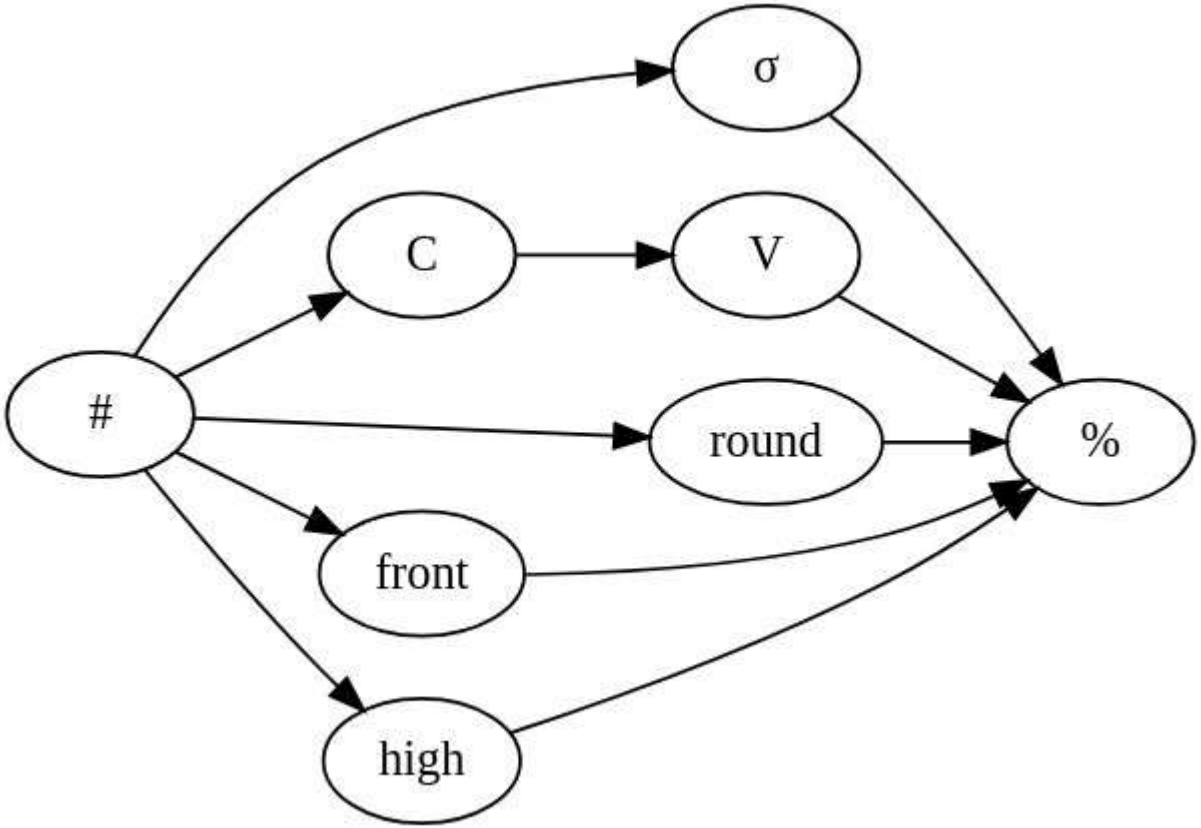
An AV puzzle (Baier, Idsardi & Lidz 2007)

Replication of Kuhl & Meltzoff, Patterson & Werker

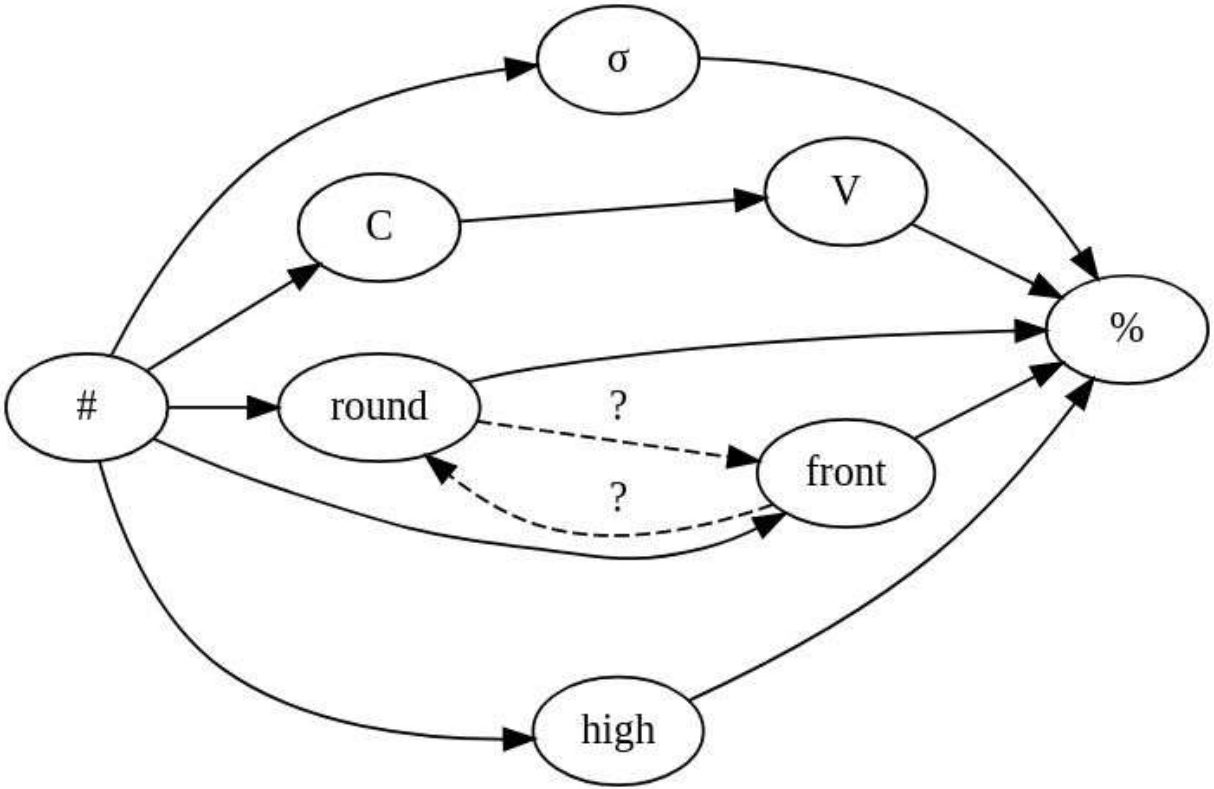
/a/, /i/, /u/, /wi/

But /wi/ ≈ /ju/ ???

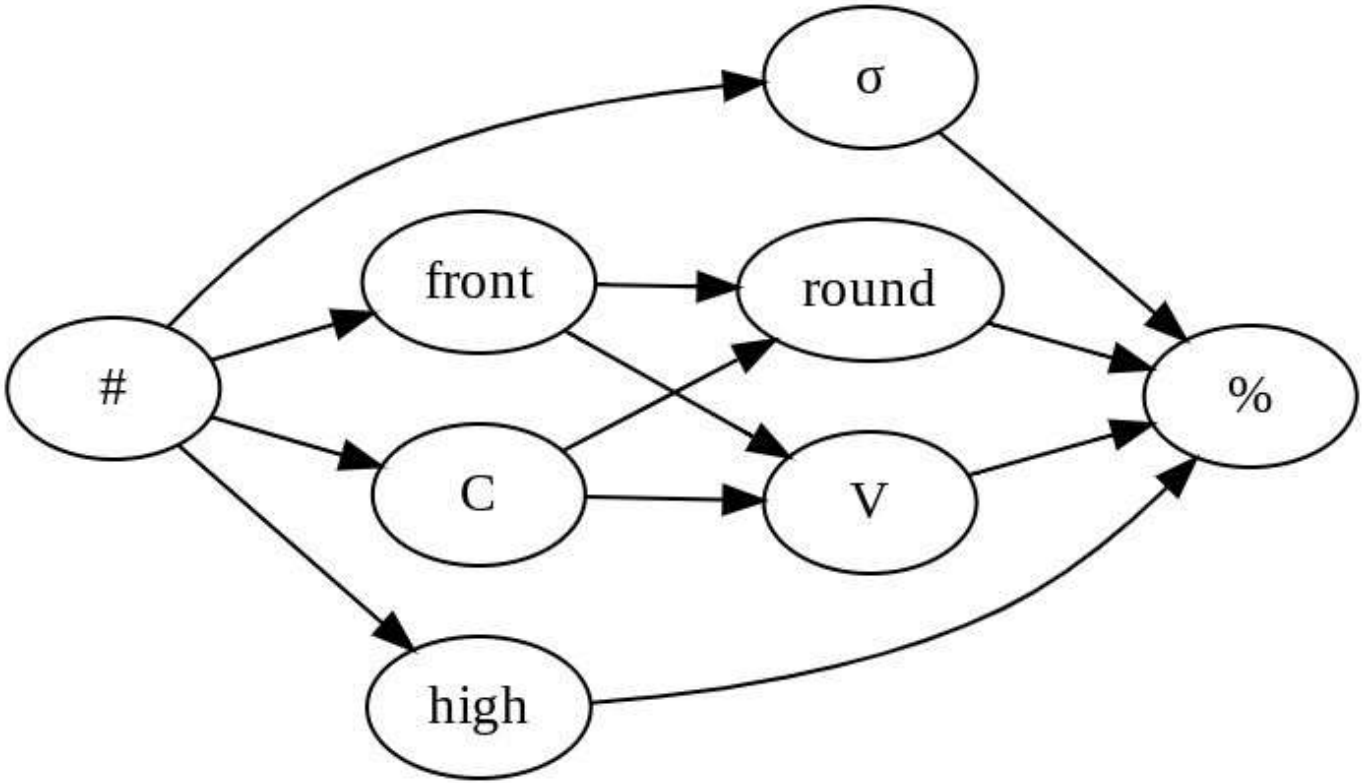
[wi] ~ [ju]



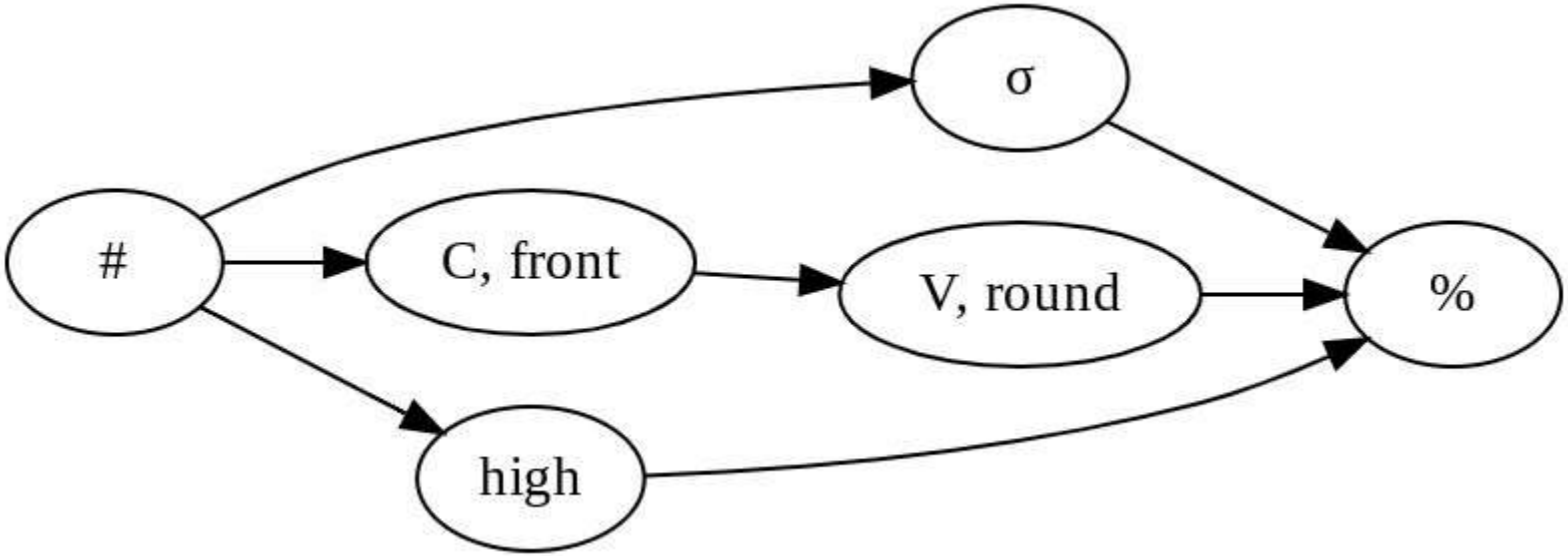
Resolving [wi] ~ [ju] (crossing the streams)



Further elaboration



Feature binding



Subsyllabic structure and proximate units

Debate about the phonological planning unit (O'Séaghdha)

Segment (English)

Syllable (Chinese, but without tone)

Mora (Japanese)

Different degrees of featural precedence and feature binding

Birdsong

From sets to ordered sets

Order in birdsong perception

Finch, canaries and budgies listening to finch songs and Schroeder complexes

Plasticity differences:

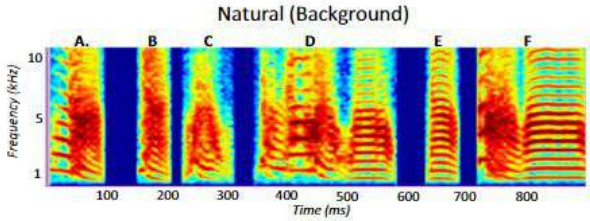
Zebra finch learn their song once as juveniles

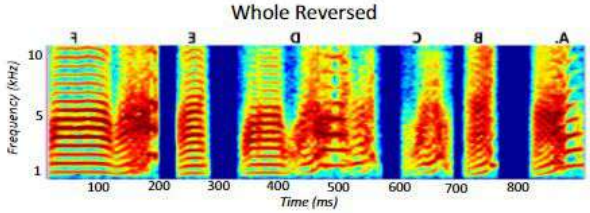
Canaries learn a new song each year during mating season

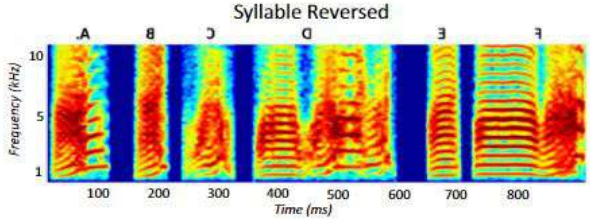
Budgies can learn new songs continually

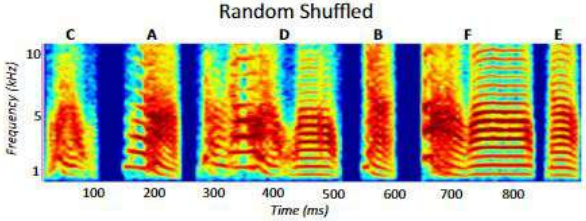
Song production is highly stereotyped

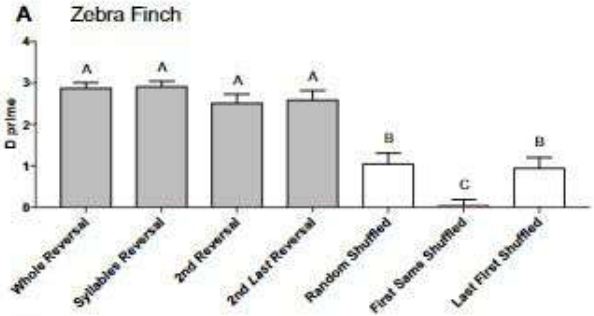
Zebra finch motif

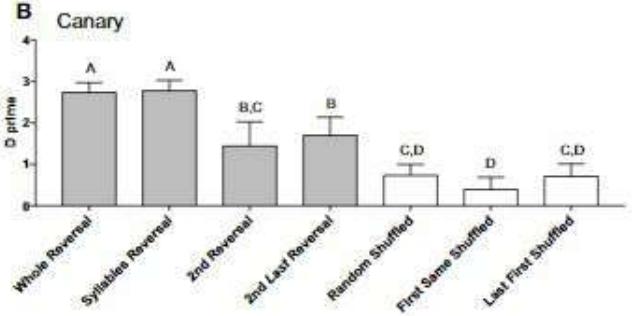


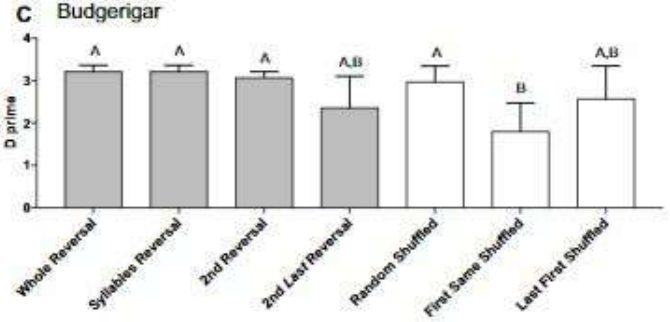




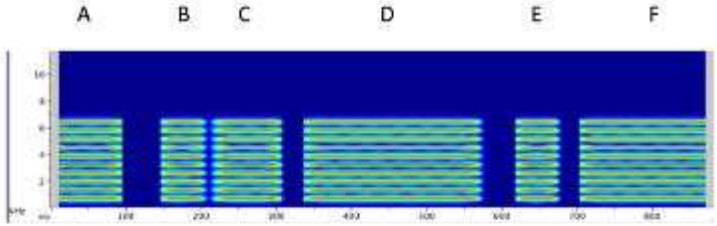




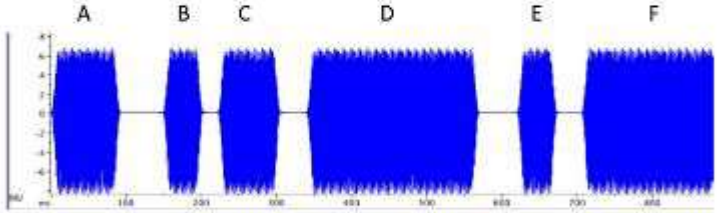




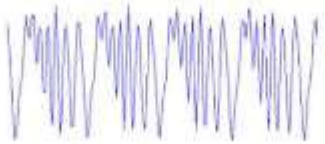
Full Schroeder Motif: Sonogram

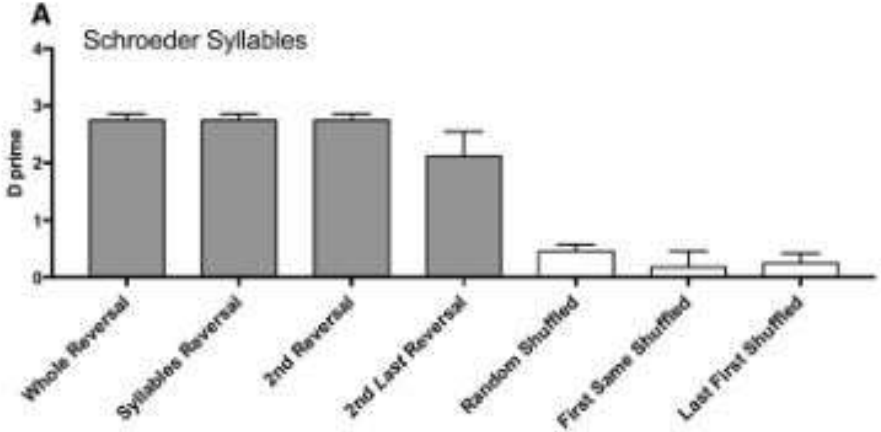


Full Schroeder Motif: Time-waveform



Several cycles of the time-waveform





Birds: interim conclusions

Finch seem to have a “grocery list” of syllables (= bag of syllables, unigrams)
{a, b, c, d, e}

Budgies do much better, do show order perception (= precedence, bigrams)
{a, b, c, d, e} + {ab, ac, ad, ae, bc, bd, be, cd, ce, de}

Can we find the neural circuit that budgies have and finch don’t?

Related to plasticity? Finch learn song once, budgies throughout life

Canaries have seasonal, hormonal plasticity (new song each year)

Sign languages

Speculations about spatial relations

Sign language symmetry

Battison: yes

Kita, van Gijn & van der Hulst: no, shared with gesture

What would speech comparisons be? Screams? Laughter? Chewing?

Sign language symmetry

Spatial symmetry (positions)

Spatio-temporal symmetry (movements)

No (pure) temporal symmetry? (Palindromes) “Able was I ere I saw Elba”

Various kinds of symmetries (perception and production aspects?)

Walking vs hopping

Rub your tummy and pat your head

Tie your shoes

Anderson 1993: differences

“Linguistic expression and its relation to modality”

Reflection: comes with symmetry calculations

Movement: (paths), several coordinate systems? (degrees of freedom)

Repetition: traverse loops more than once (“linearization”)

Gradience: ??? (Not sure that I agree here)

Simultaneity: e||f

The meanings of coincidence

<E, F, ST> with relations over spacetime?

Coincidence in spacetime for sign (contact)

Coincidence in time for spoken language (association lines)

What kinds of spatio-temporal relations are there?

Summary

Phonological representations have:

A set of events

Monadic features for each event (bag of features, unigrams)

Precedence relations between pairs of events (open bigrams, “maybe next”)

Coincidence relations between pairs of events (overlap, synchronize)

Spatial relations between pairs of events in sign

Thank you!

idsardi@umd.edu

Funding: BBI, NSF, NIH