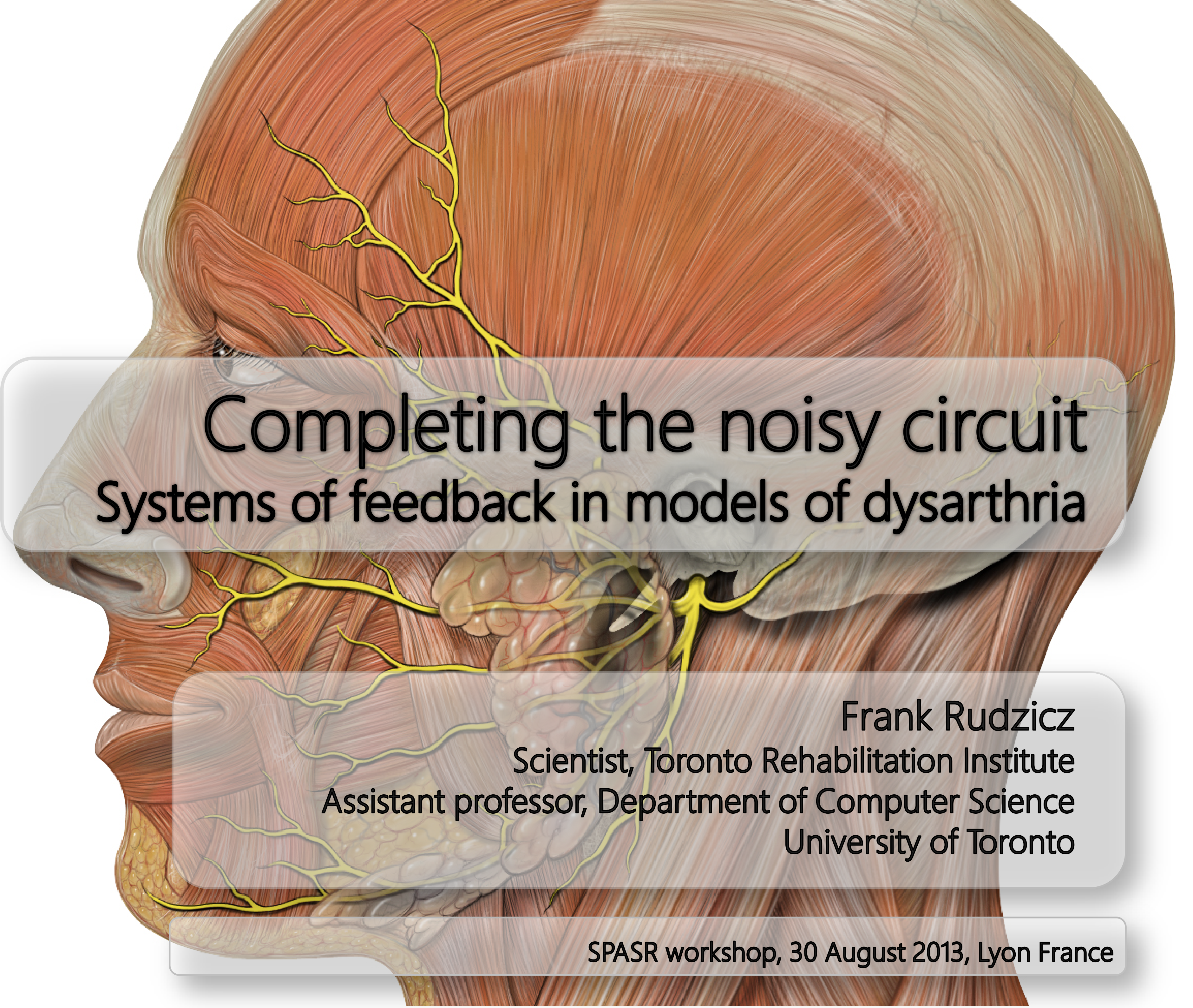


Completing the noisy circuit

Systems of feedback in models of dysarthria

Frank Rudzicz
Scientist, Toronto Rehabilitation Institute
Assistant professor, Department of Computer Science
University of Toronto

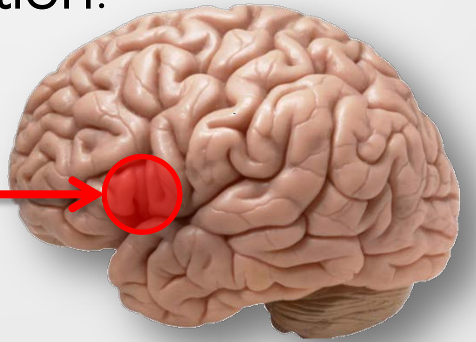
SPASR workshop, 30 August 2013, Lyon France

Studying how systems break down

- Observing how **closed systems** *fail* can be a **valuable method** in discovering how those systems **work**.

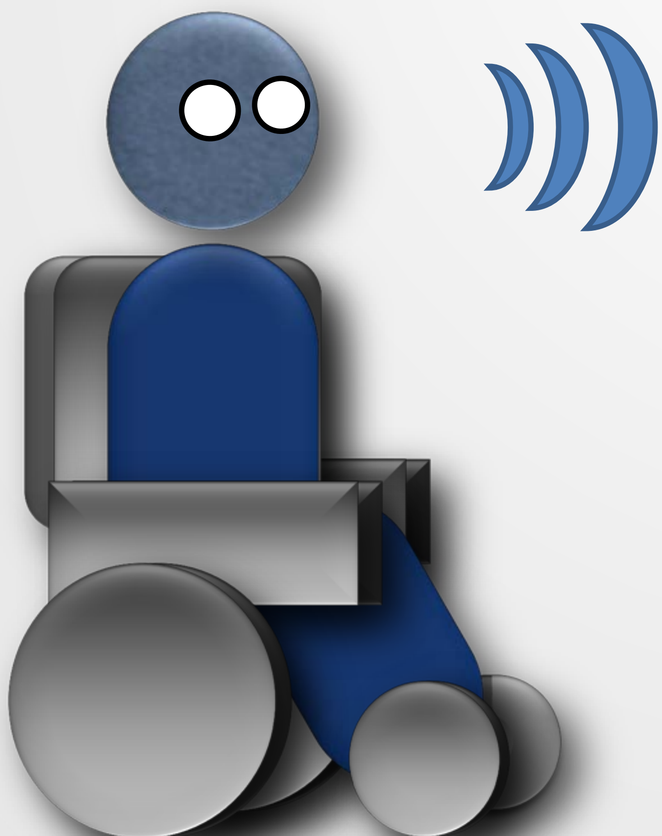
- **Paul Broca** (left) discovered, in 1861, that a **lesion** in the **left** ventro-posterior **frontal lobe** caused **expressive aphasia.**
- This was the first **direct** evidence that **language function** was **localized**.
  - It hinted at a **mechanistic** view of **speech production**.

Broca's area



Dysarthria

Neuro-motor articulatory disorders resulting in **unintelligible** speech.



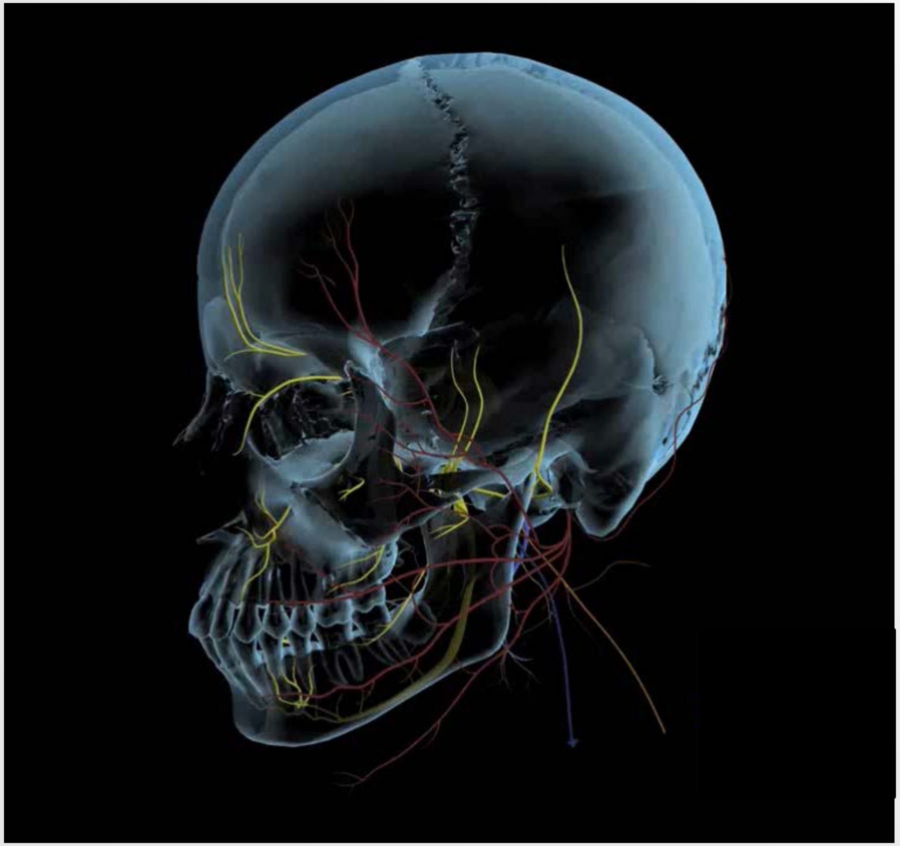
7.5 million Americans have **dysarthria**

- Cerebral palsy,
- Parkinson's,
- Amyotrophic lateral sclerosis)

(National Institute of Health)

Nosology of dysarthria

- **Types** of dysarthria are related to **specific sites** in the subcortical nervous system.

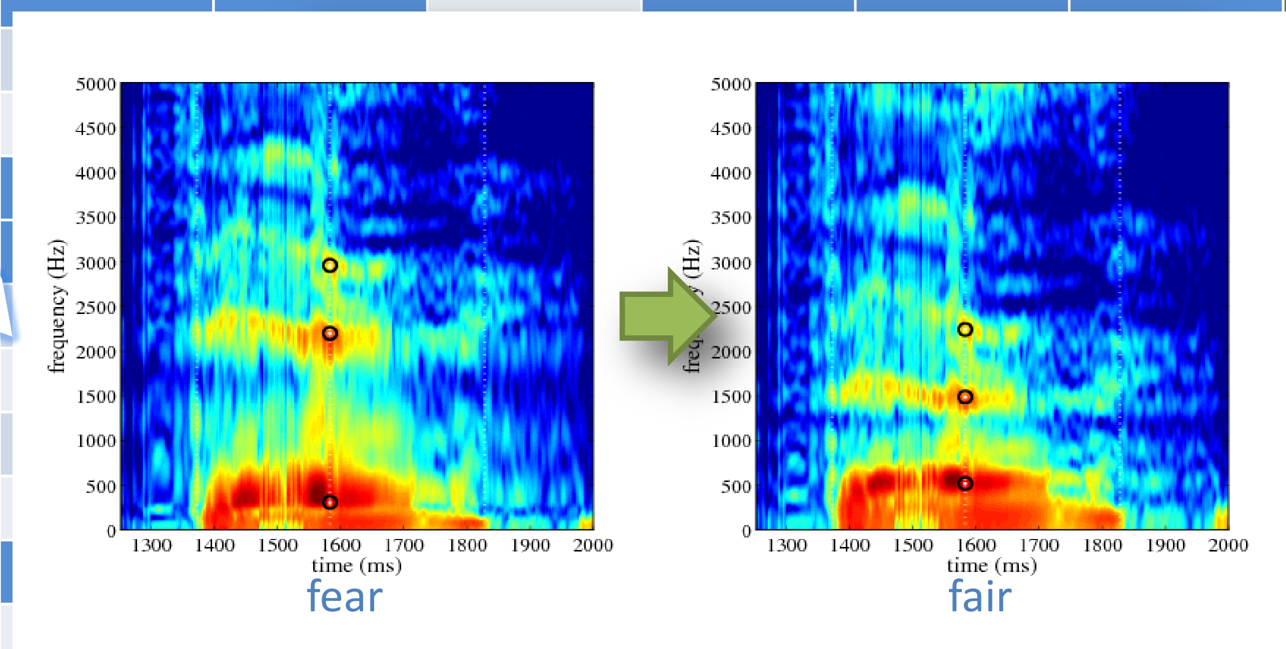


Type	Primary lesion site
Ataxic	**Cerebellum** or its outflow pathways
Flaccid	**Lower motor neuron** (≥1 cranial nerves)
Hypo-kinetic	**Basal ganglia** (esp. substantia nigra)
Hyper-kinetic	**Basal ganglia** (esp. putamen or caudate)
Spastic	**Upper motor neuron**
Spastic-flaccid	Both **upper** and **lower motor neurons**

(After Darley *et al.*, 1969)

Characteristics of dysarthria

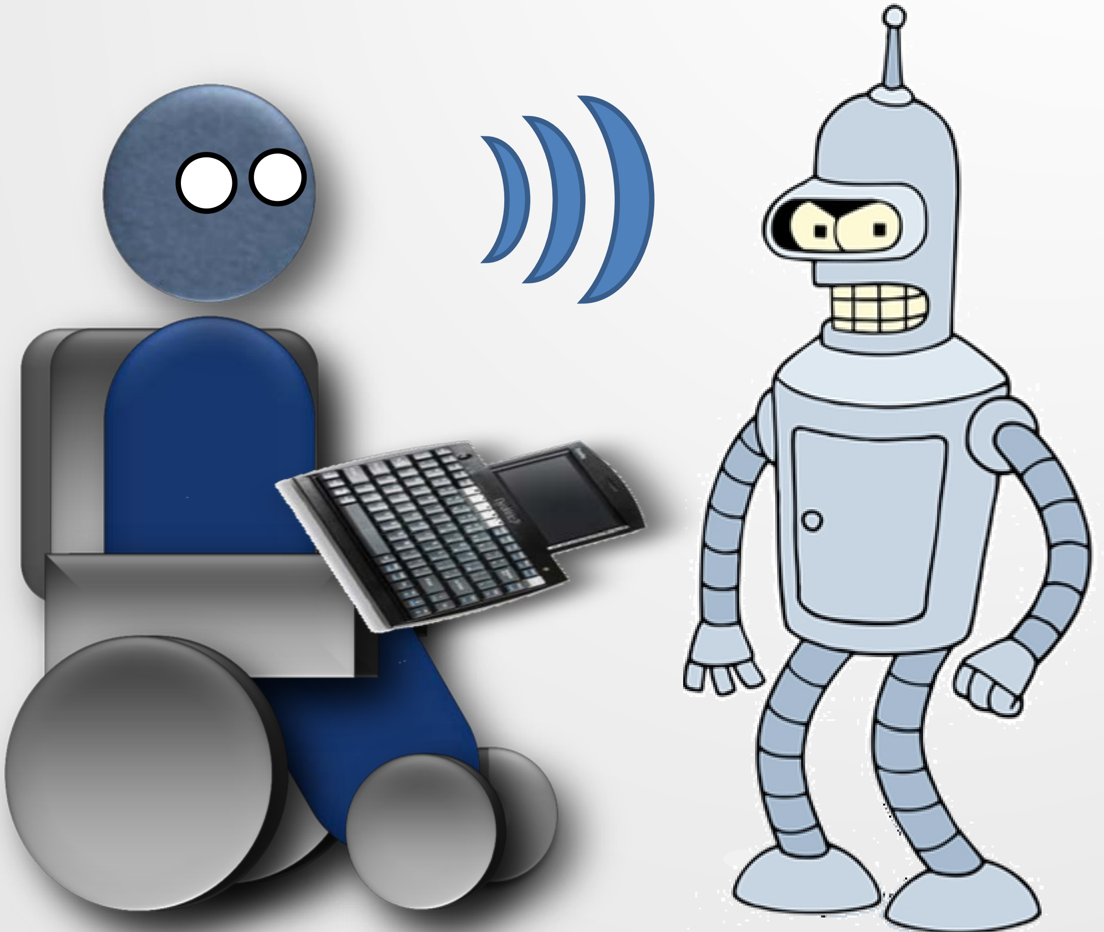
	Ataxic	Flaccid	Hypo-kinetic	Hyper-kinetic, chorea	Hyper-kinetic, dystonia	Spastic	Spastic-flaccid (ALS)
Monopitch							
Harshness							
Imprecise consonants							
Mono-loud							
Distorted vowels							
Slow rate							
Short phrases							
Hypernasal							
Prolonged intervals							
Low pitch							
Inappropriate silences							
Variable rate							
Breathy voice							
Strain-strangled voice							
…							



SPOClab signal processing and oral communication
UHN Toronto Rehabilitation Institute
University of Toronto

Dysarthria

The **broader** neuro-motor deficits associated with dysarthria can make **traditional** human-computer interaction difficult.



Can we use ASR for dysarthria?

SPOClab signal processing and oral communication

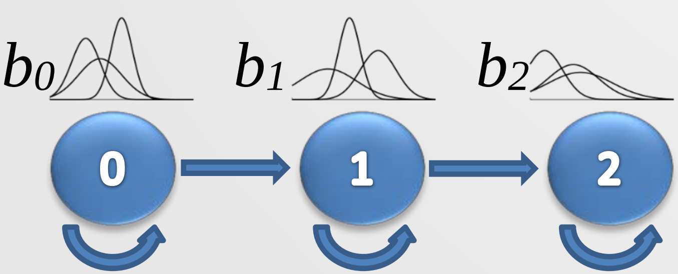
UHN Toronto Rehabilitation Institute

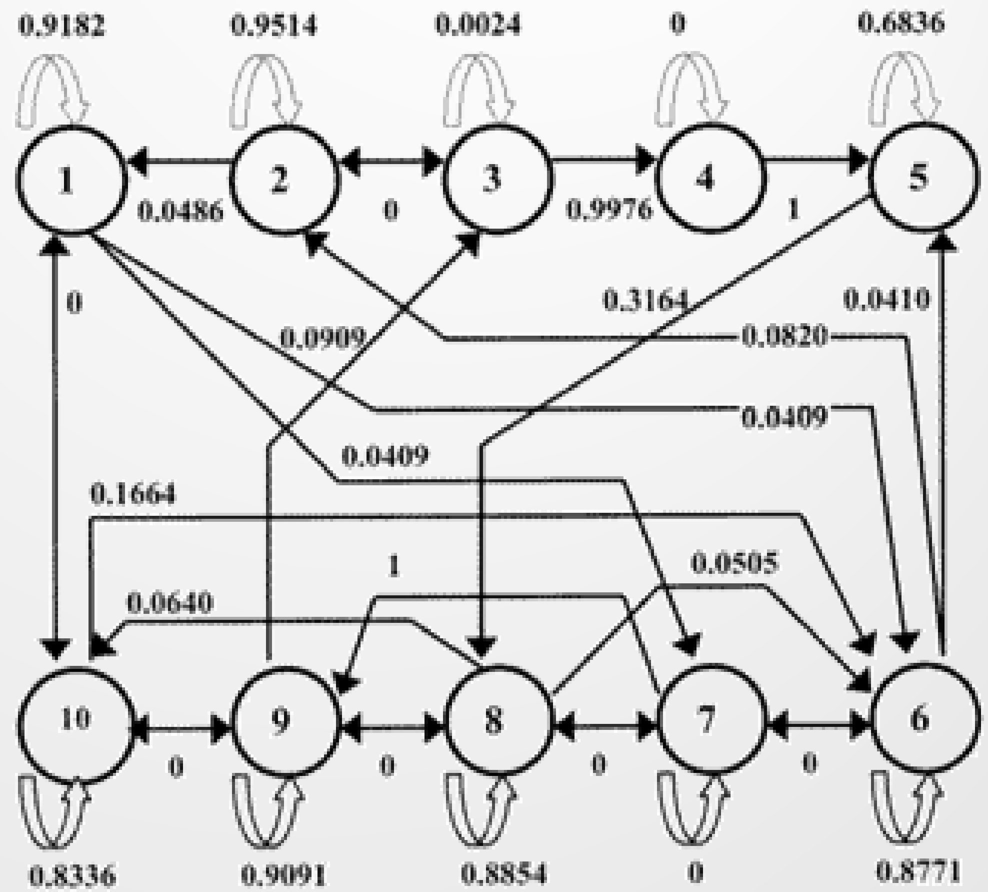
UNIVERSITY OF TORONTO

Accounting for aspects of dysarthria

- **Ergodic** HMMs can be **robust** against recurring **pauses**, and **non-speech** events.

- Polur and Miller (2005) **replaced GMM** densities **with neural networks** (after Jayaram and Abdelhamied, 1995), further **increasing accuracy**.





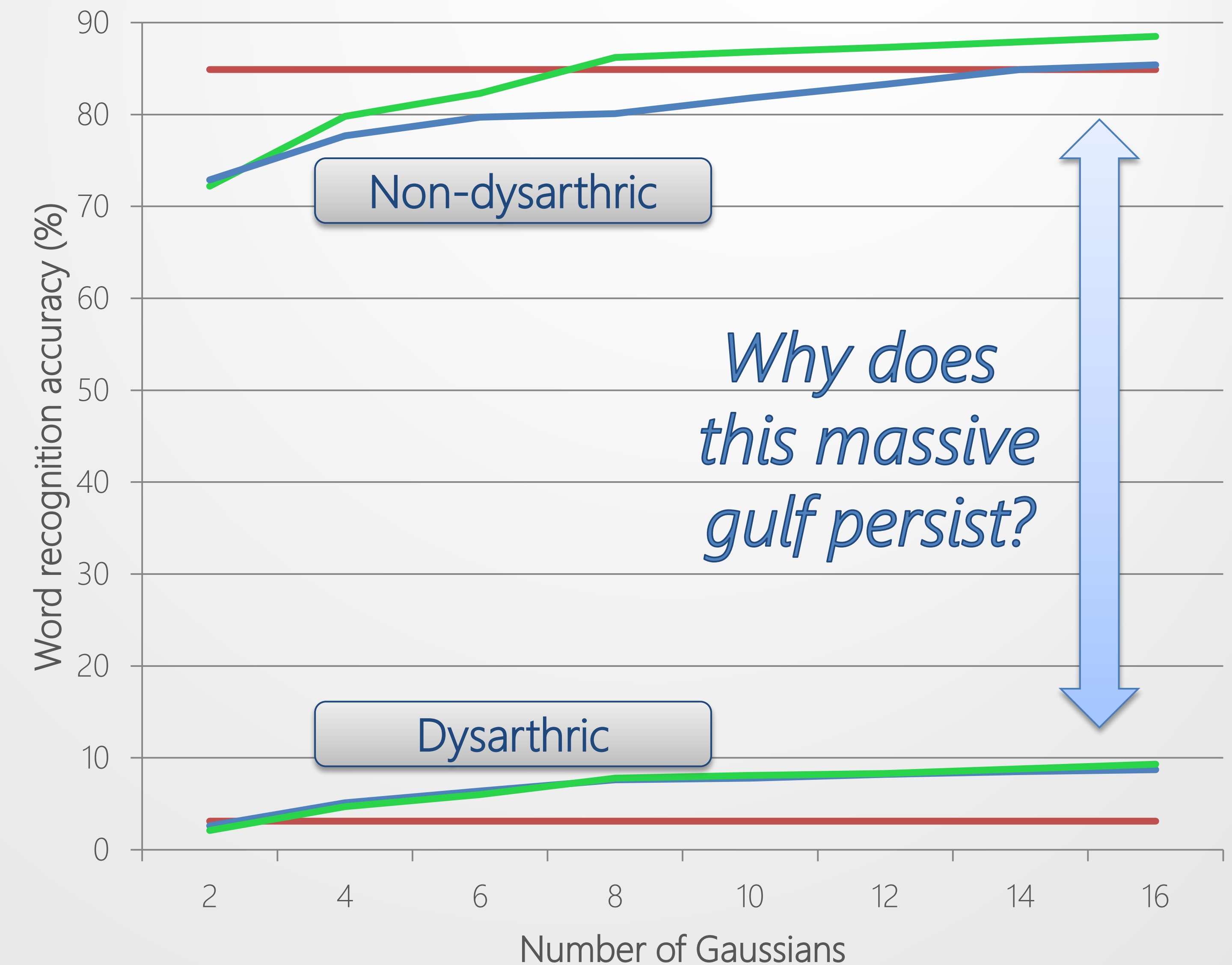
(From Polur and Miller., 2005)

Adjusting to the individual speaker

84.9%

Traditional ASR
Speaker-dependent
Speaker-retrained

3.1%

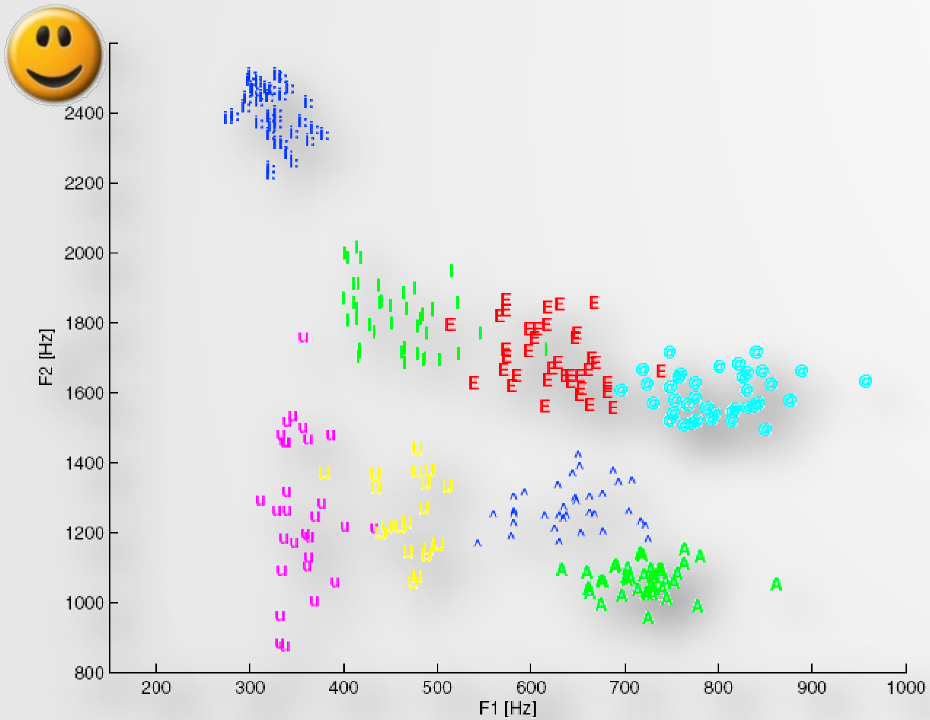


SPOClab signal processing and oral communication

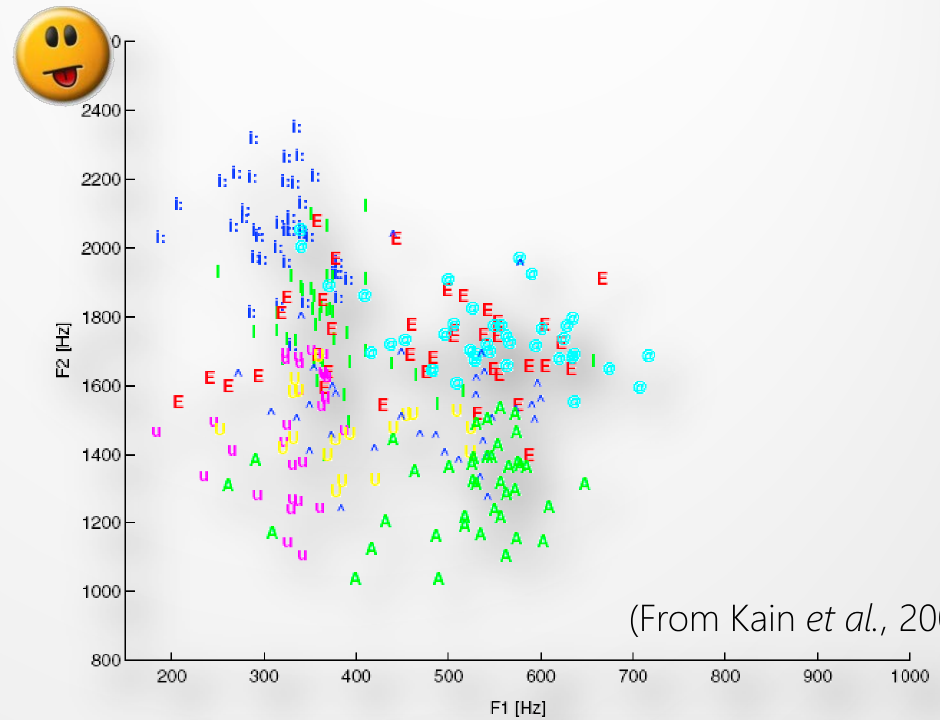
UHN Toronto Rehabilitation Institute

UNIVERSITY OF TORONTO

Acoustic ambiguity



Non-dysarthric



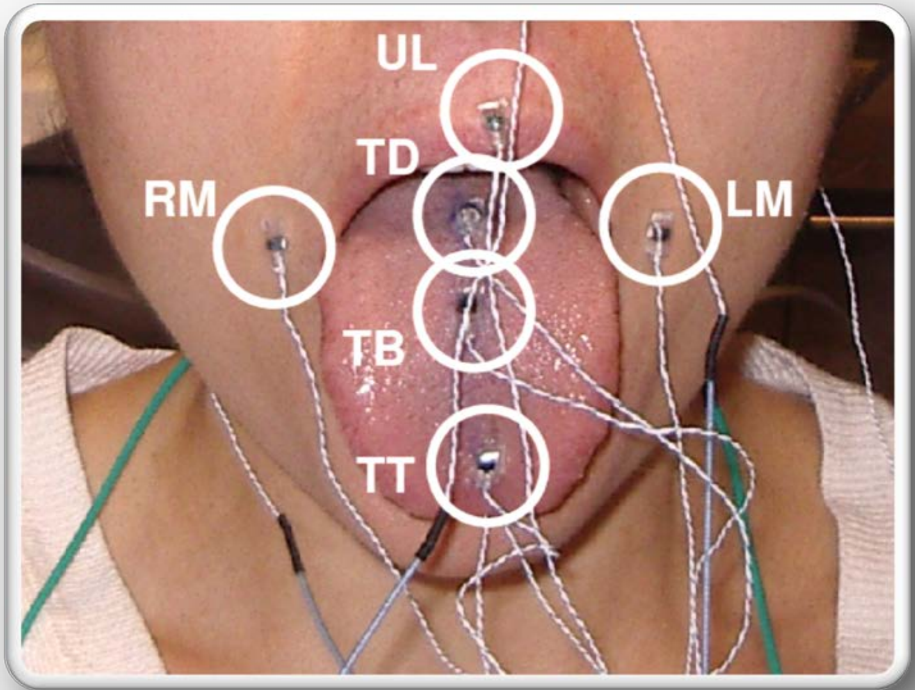
(From Kain *et al.*, 2007)

Dysarthric

This **acoustic** behaviour is indicative of underlying **articulatory** behaviour.

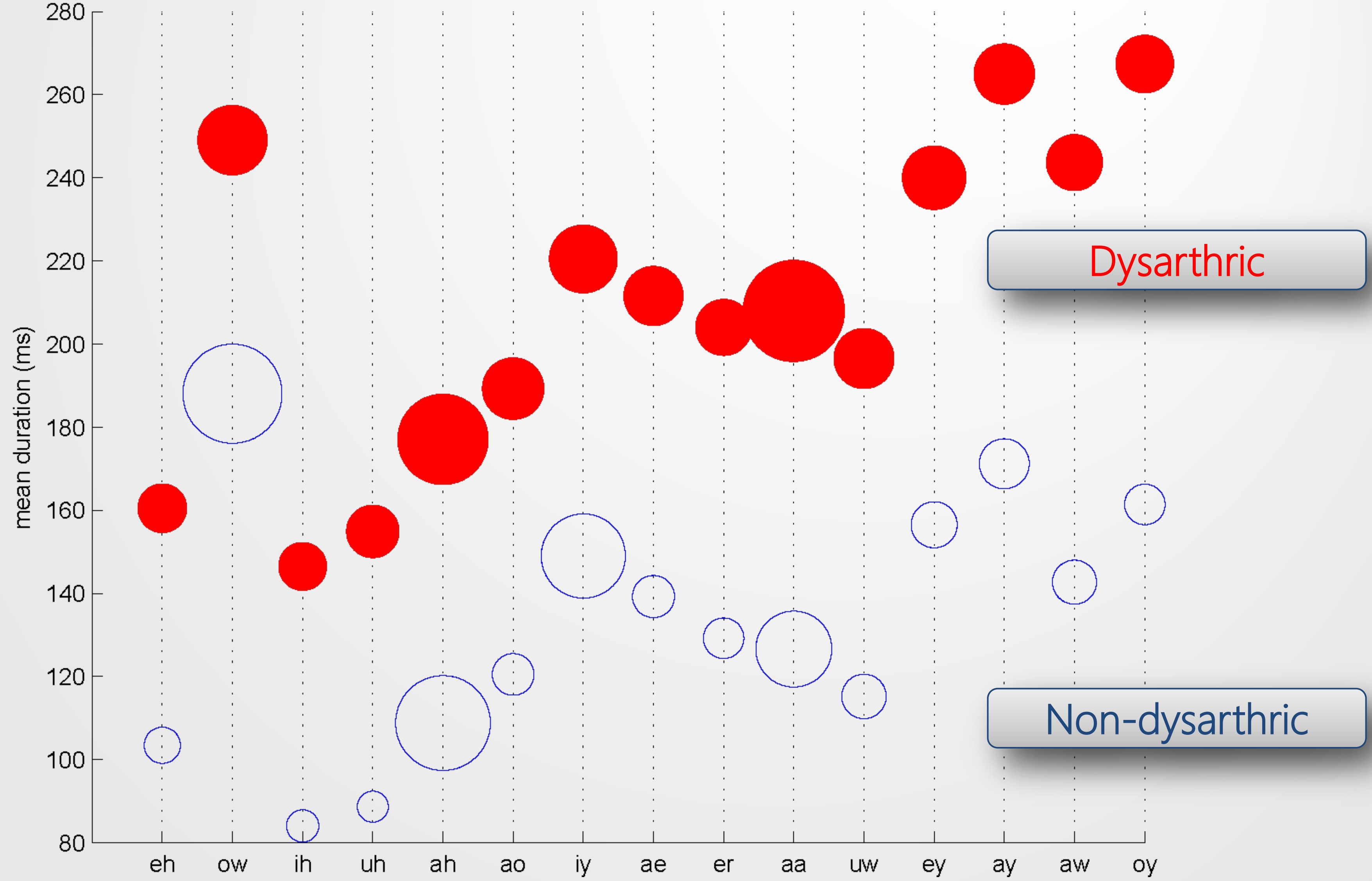
The TORGO database

- TORGO was built to train augmented ASR systems.
  - 9 subjects **with cerebral palsy**, **9 matched controls**.
  - Each reads 500—1000 prompts over **3 hours** that cover **phonemes** and **articulatory contrasts** (e.g., *meat* vs. *beat*).
  - **Electromagnetic articulography** (and video) track points to <1 mm error.



SPOClab signal processing and oral communication
UHN Toronto Rehabilitation Institute
UNIVERSITY OF TORONTO

Vowel durations in TORGO



SPOClab
signal processing and
oral communication

Information in TORGO

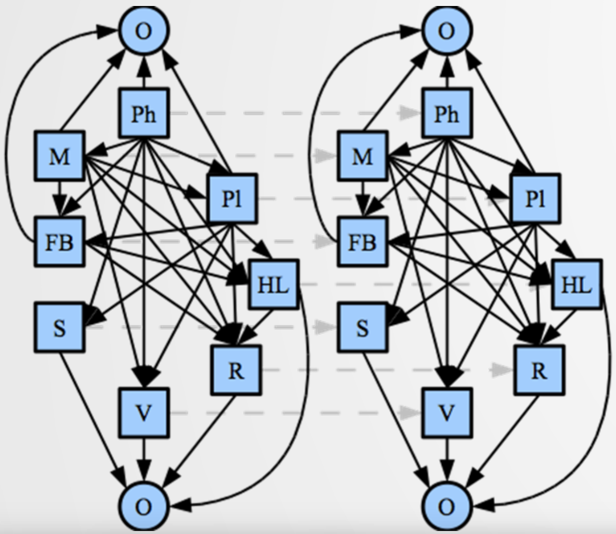
	Speaker	$H(Acous)$	$H(Artic)$	$H(Ac \mid Ar)$
Dysarthric	M01	66.37	17.16	50.30
	M04	33.36	11.31	26.25
	F03	42.38	19.33	39.47
	Average	47.34	15.93	38.68
Control	MC01	24.40	21.49	1.14
	MC03	18.63	18.34	3.93
	FC02	16.12	15.97	3.11
	Average	19.72	18.60	2.73

Dysarthric **acoustics** are far more statistic-ally disordered than the control data *but*

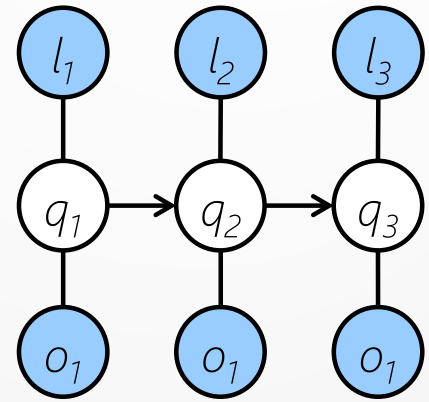
Dysarthric **articulation** is *just as* statistically ordered as the control data *yet*

Dysarthric acoustics are far less **predictable** from articulation.

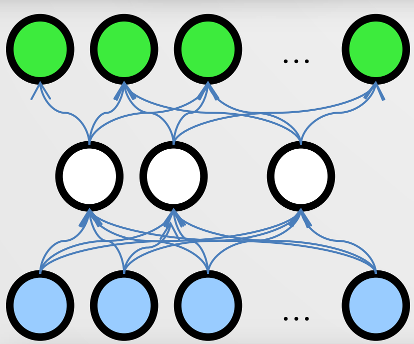
Classifying dysarthric acoustics



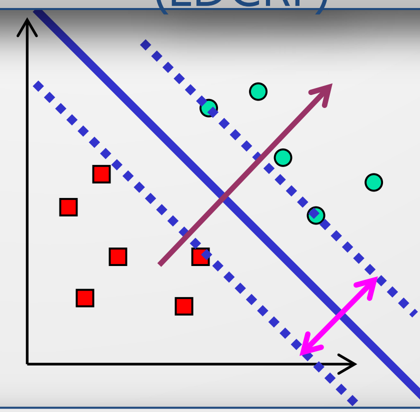
Dynamic Bayes nets
(DBN-F)



Conditional random fields
(LDCRF)

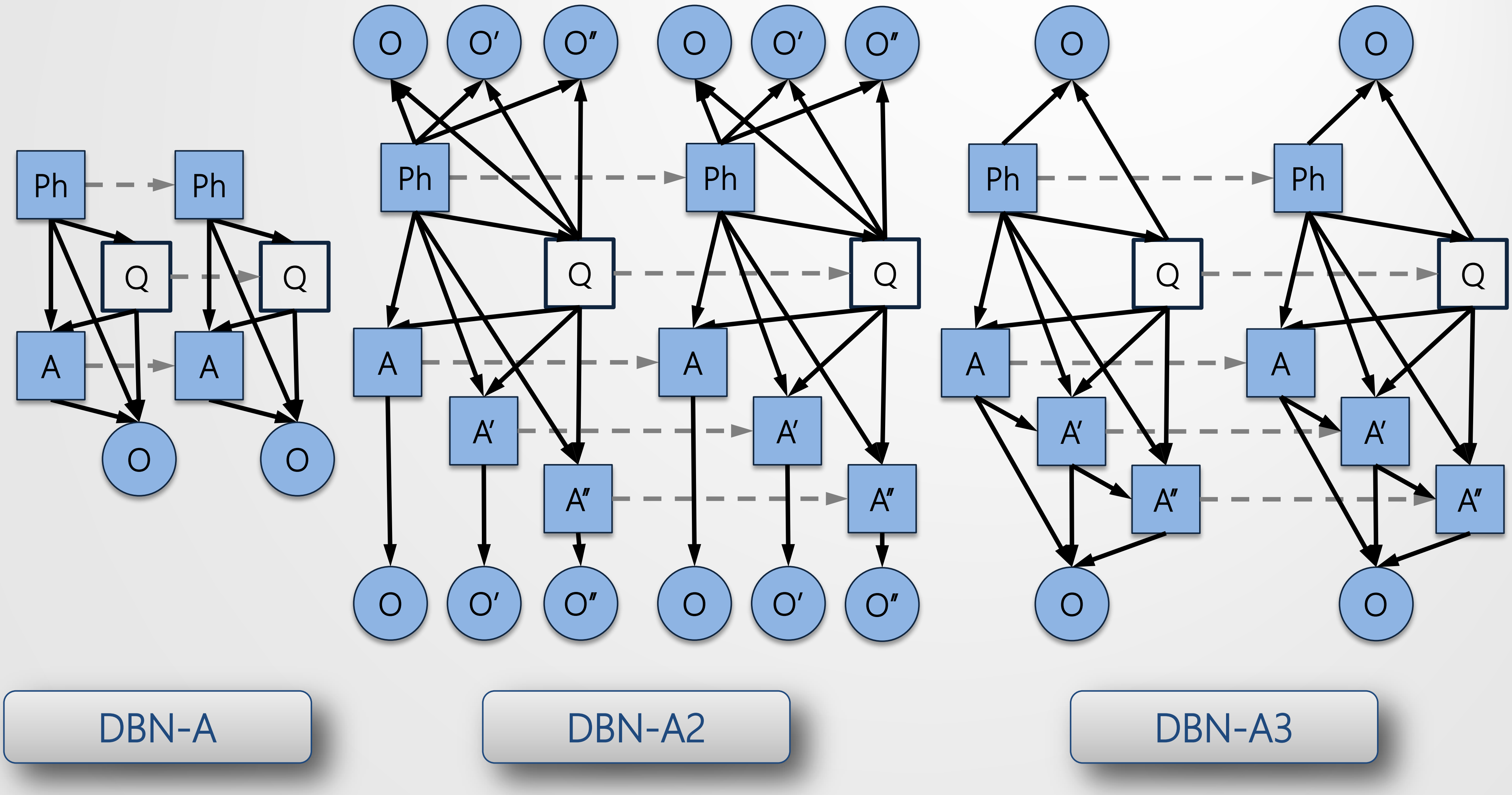


Neural networks

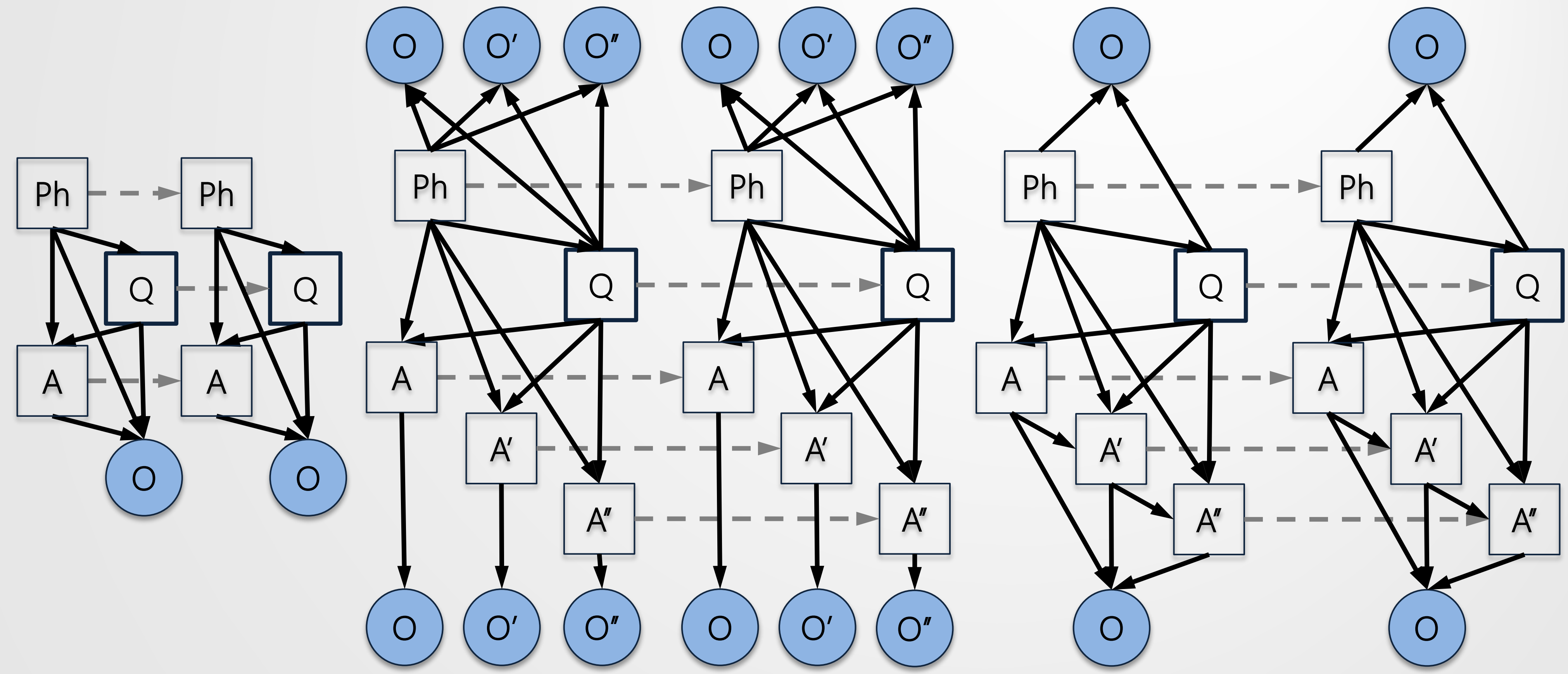


Support vector machines

Dynamic Bayes nets with EMA data



Dynamic Bayes nets with EMA data



DBN-A

DBN-A2

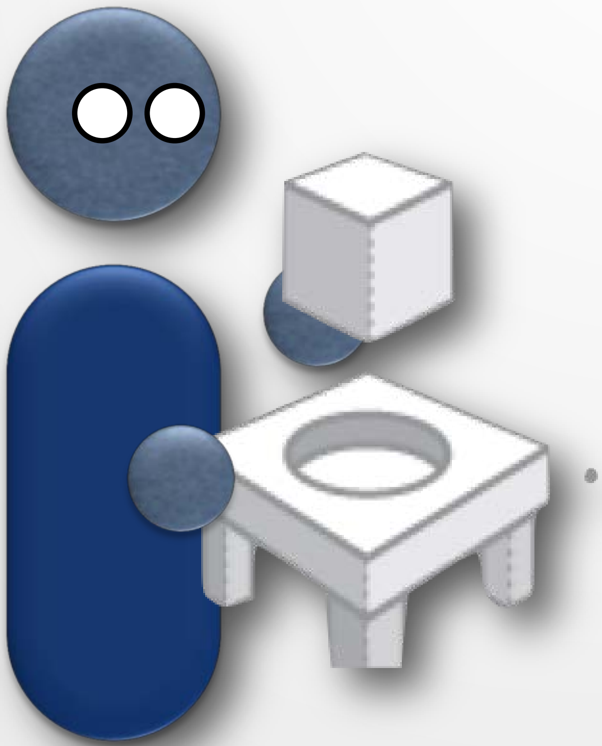
DBN-A3

Phoneme recognition

Severity of dysarthria	HMM	LDCRF	DBN		NN	
			DBN-F	DBN-A	MLP	Elman
Severe	14.1	15.2	15.0	16.4	15.5	15.6
Moderate	27.8	28.0	28.0	31.1	28.6	30.5
Mild	51.6	51.8	51.6	54.2	51.4	51.2
Control	72.8	73.5	73.3	73.6	72.6	72.7

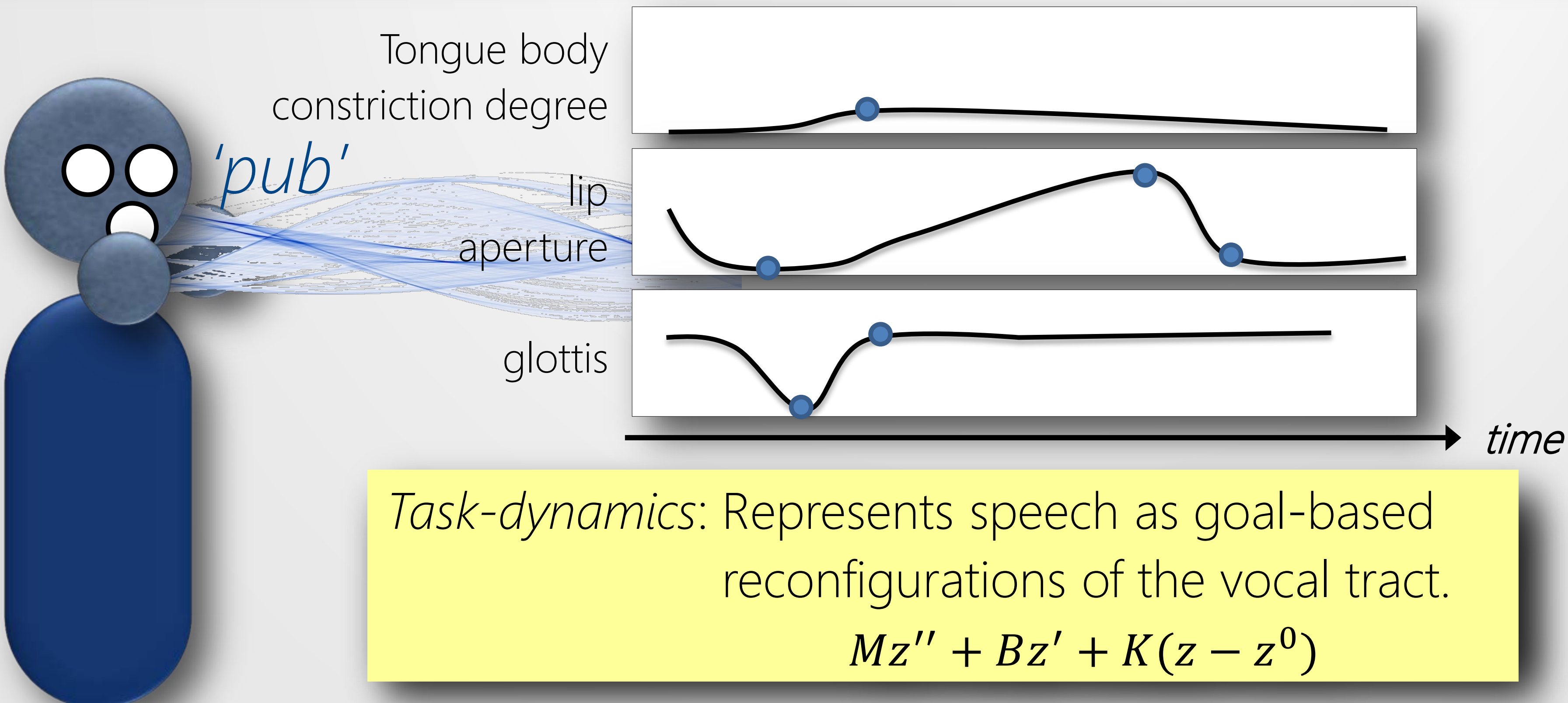
Average % phoneme accuracy (frame-level) with speaker-dependent training

Beyond discrete articulation



Dynamic speech gestures

We wish to classify dysarthric speech in a low-dimensional and informative space that incorporates **goal-based** and **long-term dynamics.**



Task-dynamics: Represents speech as goal-based reconfigurations of the vocal tract.

$$Mz'' + Bz' + K(z - z^0)$$

Characteristics of dysarthria

	Ataxic	Flaccid	Hypo-kinetic	Hyper-kinetic, chorea	Hyper-kinetic, dystonia	Spastic	Spastic-flaccid (ALS)
Monopitch							
Harshness							
Imprecise consonants							
Mono-loud							
Distorted vowels							
Slow rate							
Short phrases							
Hypernasal							
Prolonged intervals							
Low pitch							
Inappropriate silences							
Variable rate							
Breathy voice							
Strain-strangled voice							
…							

Task-dynamics:

$$Mz'' + Bz' + K(z - z^0)$$

Aspects to consider

- As we develop an **extension** or **alternative** to task dynamics, we have to consider:

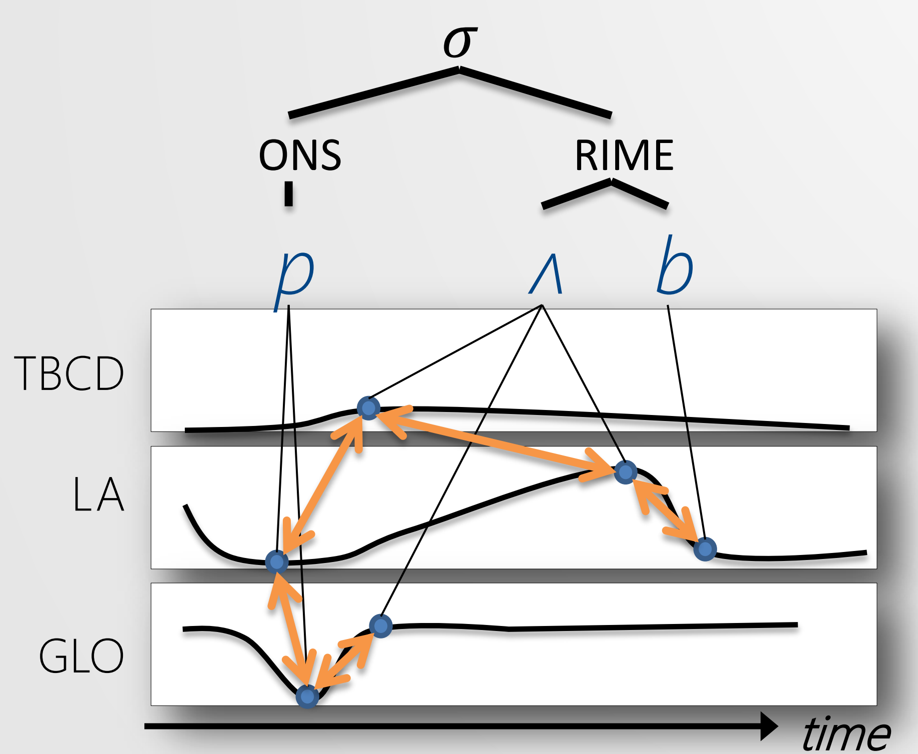
1. **Timing.**
   a) Inter-articulator co-ordination.
   b) Rhythm.

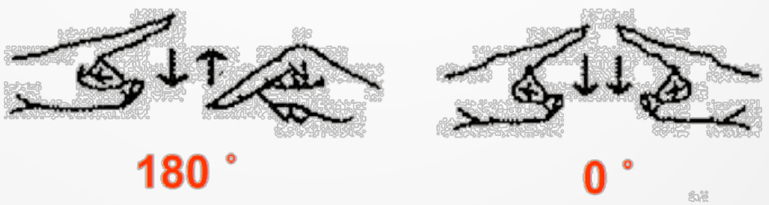
2. **Feedback.**
   a) Acoustic, proprioceptive, and tactile.

3. **Higher-level features**
   a) Syntax and meaning

1. Timing

- In TD, **pairs of goals** are **dynamically coupled** in time.
- Articulators are **phase-locked** (0˚ or 180˚; Goldstein *et al.*, 2005)





- **(C)CV** pairs stabilize **in-phase**.
- **V(C)C** pairs stabilize **anti-phase**.
- Kinematic errors occur when competing gestures are **repeated** and tend to stabilize incorrectly.
  - e.g., repeat *koptop* (Nam *et al*, 2010).

SPOClab signal processing and oral communication

UHN Toronto Rehabilitation Institute

UNIVERSITY OF TORONTO

1. Timing

- Cerebellar **ataxia** often **prohibits** control over more than one articulator at a time.
  - **Apraxia** generates incorrect motor **plans**, wholly **distorting** gestural **goals**, hence timing.

- **Dysarthric** speech **nearly equally** consists of **steady-states** (49.95%) and **transitions** (50.05%) (Vollmer, 1997).
  - **Typical** speech consists of ~**82.14%** steady-states.

1. Timing/rhythm

- **Rhythm** (the distribution of **emphasis**) is *not* part of TD.

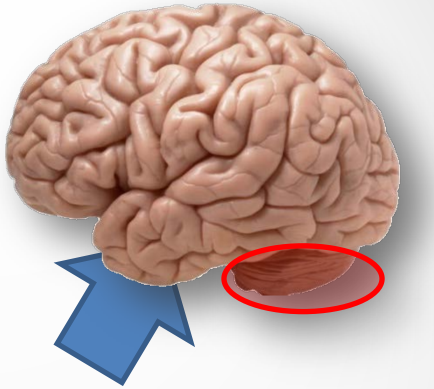
- **Tremor** behaves as oscillations about an equilibrium.
  - There is **evidence** that people with **Parkinson's** coordinate **voluntary** movement with **involuntary** tremors (Kent *et al.*, 2000).

- **Rhythm** in **ataxic** dysarthria formalized by aberrations in a 'scanning index', SI, consisting of syllable lengths S_i,

$$SI = \frac{\prod_{i=1}^{n} S_i}{\left(\frac{\sum_{i=1}^{n} S_i}{n}\right)^n}$$

(Ackermann and Hertrich, 1994))

2. Feedback



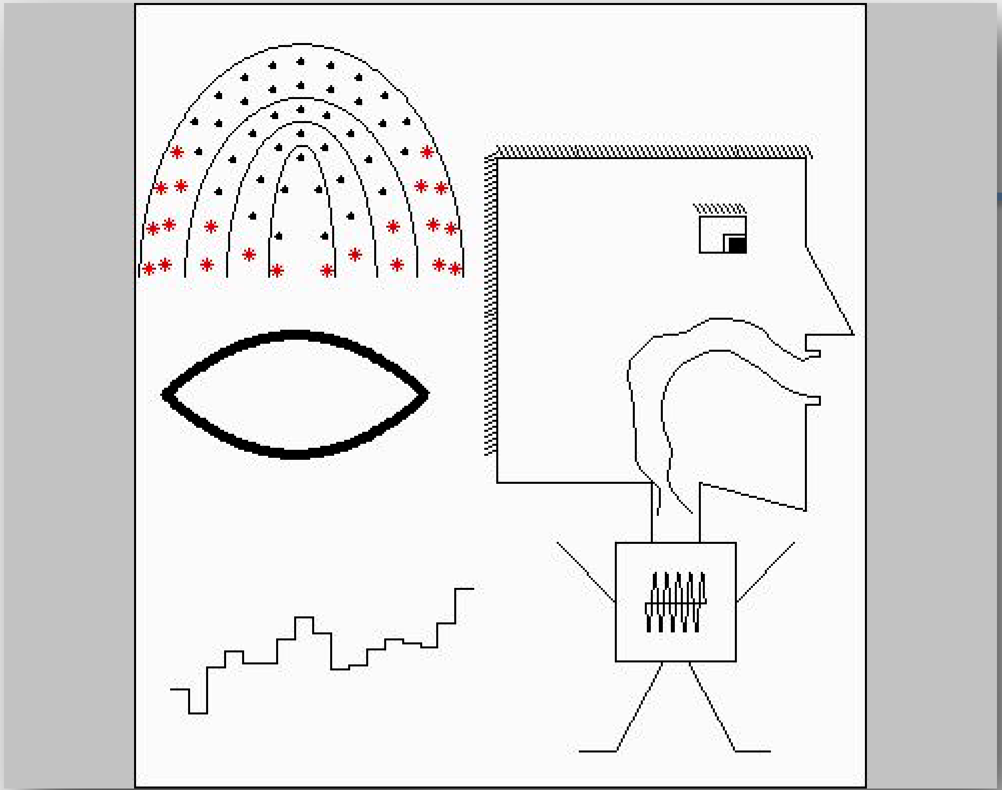
- Dysarthria can affect **sensory** cranial nerves.

- **Parkinson's disease** reduces **temporal** discrimination in **tactile**, **auditory**, and **visual** stimuli.
  - Likely explanation is that **damage** to the basal ganglia **prohibits** the formation of sensory targets (Kent *et al.*, 2000).
  - The result is **underestimated** movement.

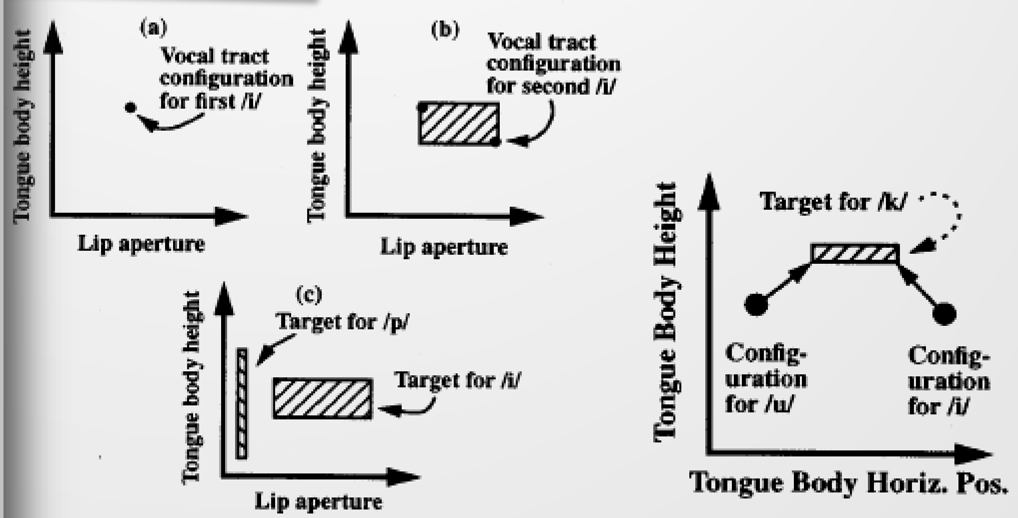
- **Cerebellar disease** results in **dysmetria** since the **internal model** of the **skeletomuscular system** is **dysfunctional**.
  - The cerebellum is apparently used in the preparation and revision of movements.

2. Feedback and DIVA

- The DIVA model is **supposed** to model feedback, but is largely **speculative** on **neurological** aspects.
- Here, **sound targets** and **somatosensory targets** are **learned** during 'babbling' and **modify** articulatory **goals**.



- This is meant to imitate the cerebellum (or basal ganglia).



3. Further into the brain with aphasia

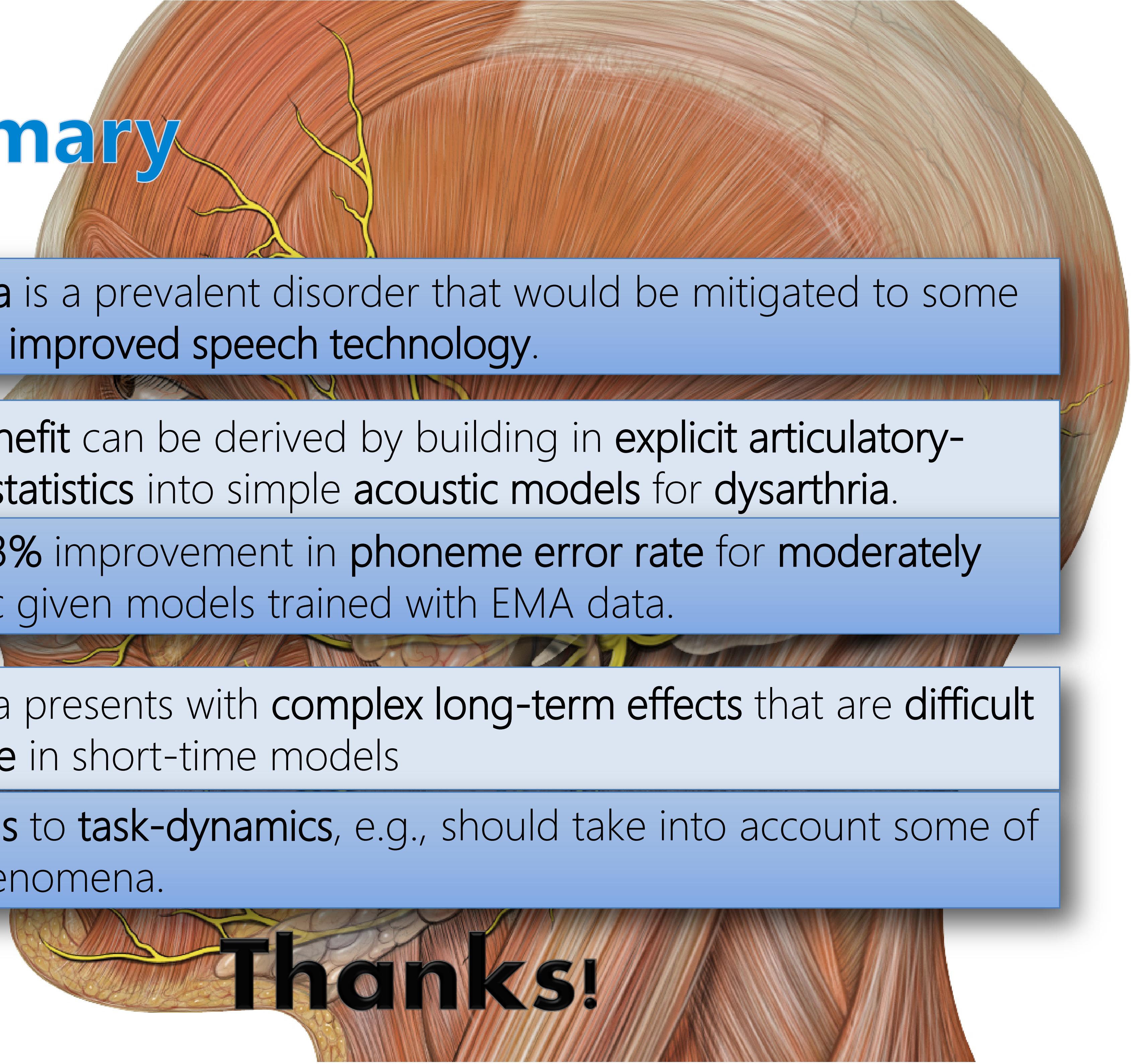
Broca's aphasia

- **Reduced** hierarchical **syntax**.
- **Anomia**.
- **Reduced** "mirroring" between **observation** and **execution** of **gestures** (Rizzolatti & Arbib, 1998).

Wernicke's aphasia

- **Normal** intonation/rhythm.
- **Meaningless** words.
- '**Jumbled**' syntax.
- **Reduced** comprehension.

Summary



Dysarthria is a prevalent disorder that would be mitigated to some extent by **improved speech technology**.

Some **benefit** can be derived by building in **explicit articulatory-acoustic statistics** into simple **acoustic models** for **dysarthria**.

About **3.3%** improvement in **phoneme error rate** for **moderately** dysarthric given models trained with EMA data.

Dysarthria presents with **complex long-term effects** that are **difficult to capture** in short-time models

Extensions to **task-dynamics**, e.g., should take into account some of these phenomena.

Thanks!