



# Intentionality in Speech

## *Implications for Computational Models*

***Prof. Roger K. Moore***

Chair of Spoken Language Processing  
Dept. Computer Science, University of Sheffield, UK  
*(Visiting Prof., Dept. Phonetics, University College London)*  
*(Visiting Prof., Bristol Robotics Lab.)*





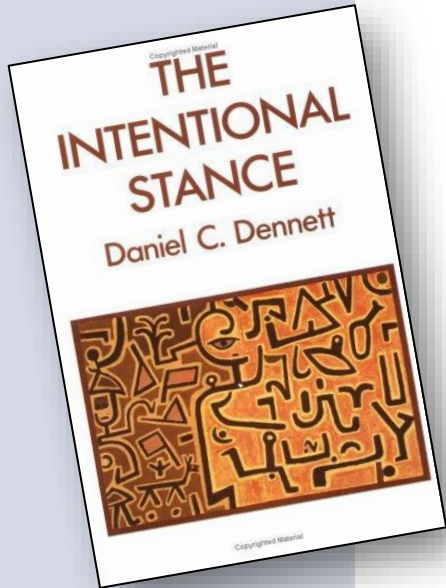
# Intentional Speech Implications of Personal Models

**Prof. Andrew Senior**



Chair of Speech Processing  
Dept. Computer Science  
(Visiting Prof., Dept. Computer Science, City College London)  
(Visiting Prof., Bristol Robotics Lab.)

# Teleological Behaviour



Dennett, D. (1989).  
*The Intentional  
Stance*. MIT Press.

- The behaviour of (*intelligent*) living systems is **intentional!**
- This doesn't mean that an organism 'knows' what it is doing!
- It simply means that an organism has **preferred states**, and that actions are selected in order to achieve those states
- This places a focus, not on actions, but on the **consequences** of actions
- This, in turn, leads to very interesting forms of **coupling** between ...
  - an agent and its environment
  - an agent and another agent

# Communicating Intentions

- Signalling involves physical/mental effort
- Large effort creates clear signals but uses more energy (*and vice versa*)
- The 'target' is a perception not a signal
- So optimisation is over competing perceptions not competing signals
- The intention is sufficient **contrast** at the pragmatic level (*leading to suitable compensations at the semantic, syntactic, lexical, phonemic, phonetic and acoustic levels*)
- The **obstacles** are ...
  - alternative interpretations (*internal*)
  - competing signals (*external*)



# Communicating Intentions

“I ... do ... not ... know”

“ I do not know”

“I don't know”

“I dunno”

“dunno”

☹️ ãããã 🌀

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405.

- Signalling involves physical/mental effort
- Large effort creates clear signals but uses more energy (*and vice versa*)
- The ‘target’ is a perception not a signal
- So optimisation is over competing perceptions not competing signals
- The intention is sufficient **contrast** at the pragmatic level (*leading to suitable compensations at the semantic, syntactic, lexical, phonemic, phonetic and acoustic levels*)
- The **obstacles** are ...
  - alternative interpretations (*internal*)
  - competing signals (*external*)

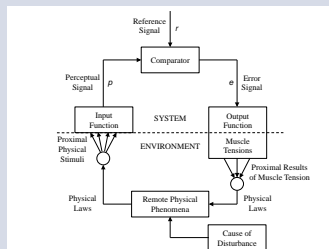


# Feedback



- The structural coupling of an agent with its environment (*including other agents*) implies **feedback**
- Feedback is a **regulatory** process
- Feedback facilitates ...
  - the management of energy and entropy
  - the maintenance of stability
  - the comparison of achievements against intentions

Perceptual  
Control  
Theory



*“feedback ... is the central and determining factor in all observed behavior”*

**W. T. Powers (1973). *Behaviour: The Control of Perception*, Aldine, Chicago.**

# Evidence for Such Behaviour

- People naturally tend to speak louder/differently in noise (*Lombard, 1911*)
- Caregivers talk differently to children (*Fernald, 1985*)
- Speakers actively control articulatory effort (*Lindblom, 1990*)
- Users talk differently to machines (*Moore & Morris, 1992*)
- Being able to hear your own voice has a profound effect on speaking (*as evidenced by the need for sidetone on a telephone*)
- Hearing-impaired individuals can have great difficulty maintaining clear pronunciations (*or level control*)
- Delayed auditory feedback causes stuttering-like behaviour
- People with speaking difficulties (*e.g. caused by cerebral palsy*) report that it takes immense effort to produce even the simplest utterance
- Altered auditory feedback evokes compensations (*Munhall et al, 2009; MacDonald et al, 2011*)



# Consequences for SLP

- Need computational paradigms that are able to accommodate such dependencies
- Communicative **obstacles** are overcome using ...
  - sufficient effort
  - feedback
- Communicative **effort** is related to ...
  - the fidelity of the representations
  - the depth of the searches



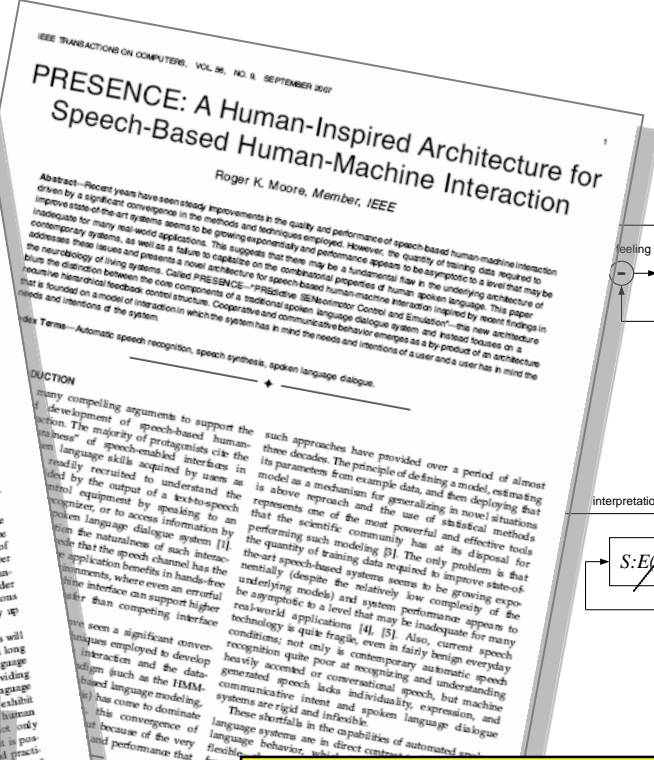


# PreSenCE

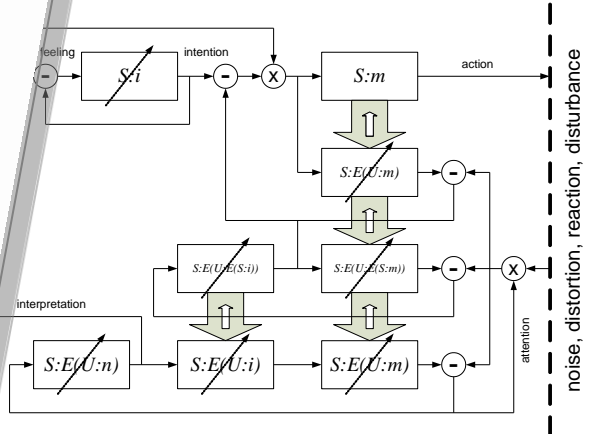
## Predictive Sensorimotor Control and Emulation



**Moore, R. K. (2007). Spoken language processing: piecing together the puzzle. *Speech Communication*, 49, 418-435.**

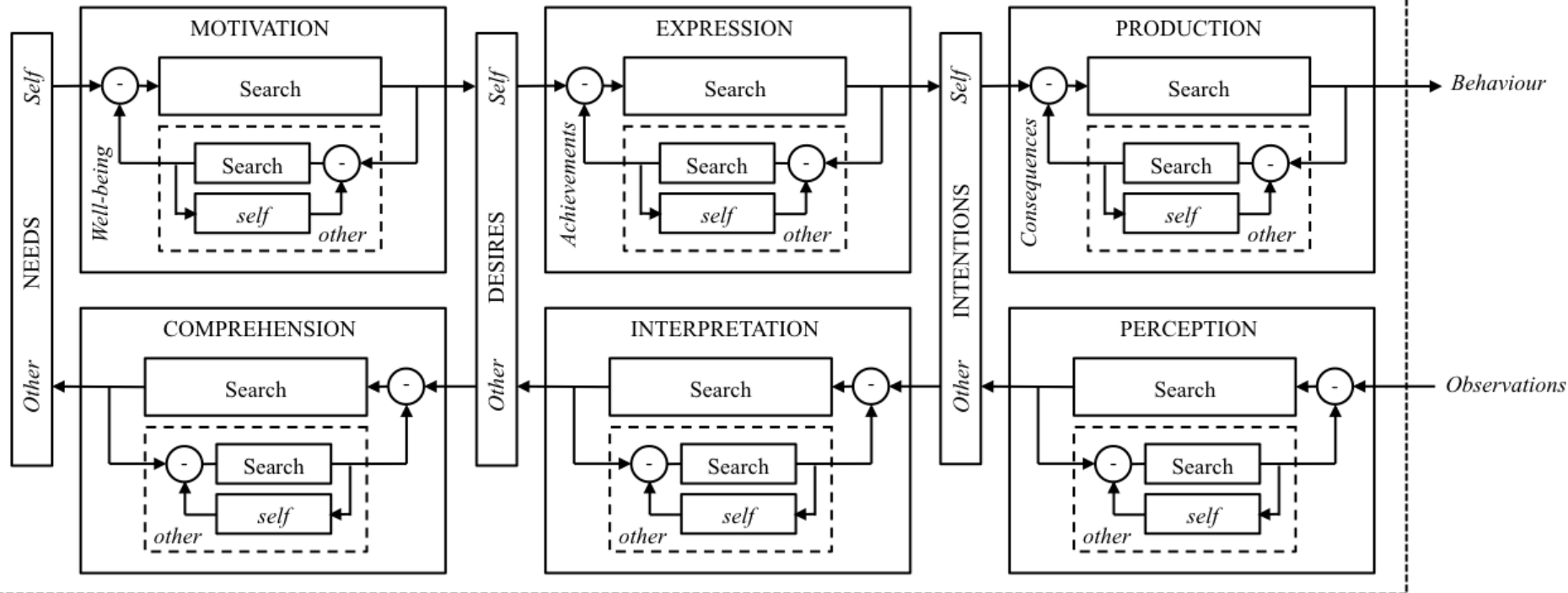


**Moore, R. K. (2007). PRESENCE: A human-inspired architecture for speech-based human-machine interaction. *IEEE Trans. Computers*, 56(9), 1176-1188.**



# Needs-Driven MBDIAC Agent

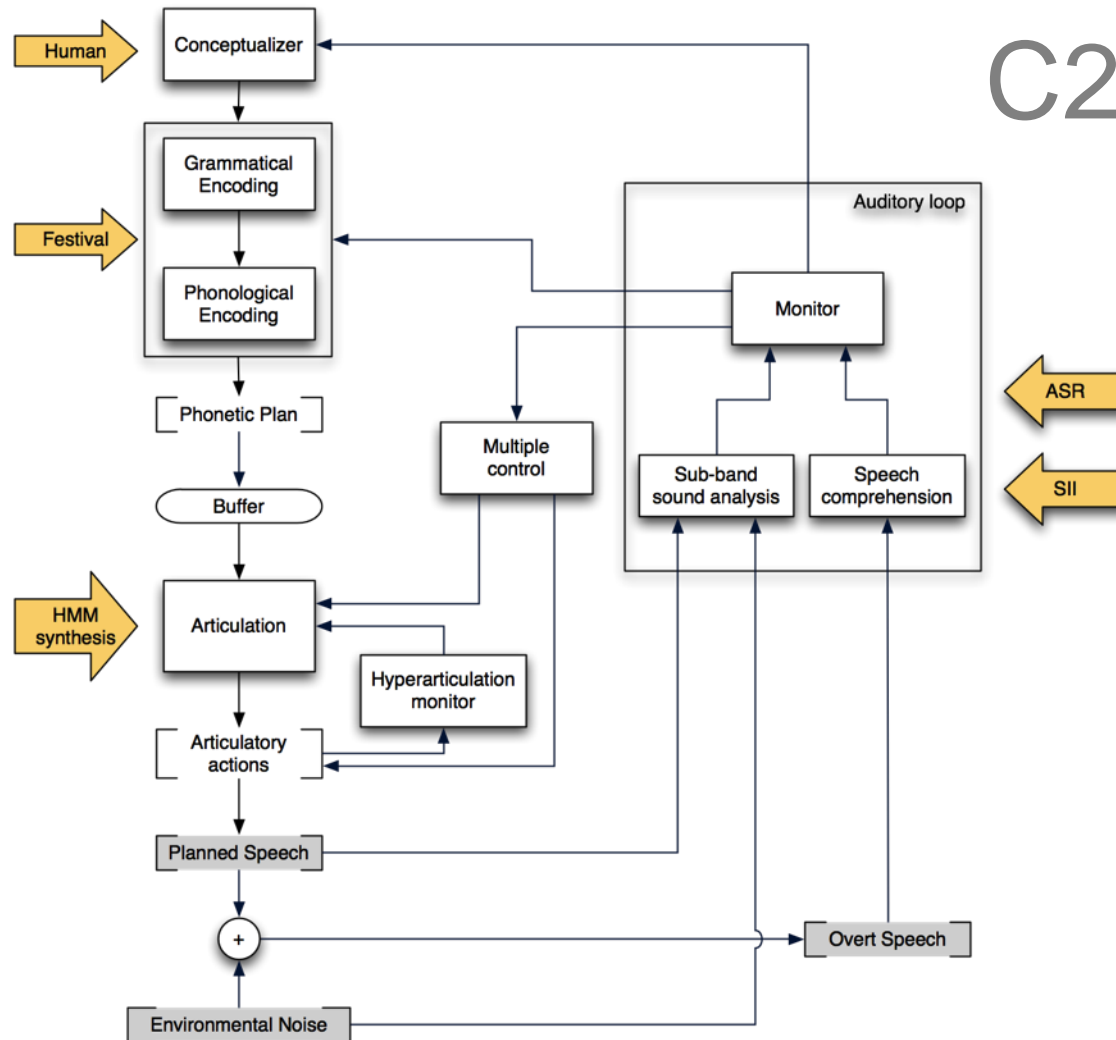
## NEEDS-DRIVEN COMMUNICATIVE AGENT



*Mutual Beliefs* *Desires* *Intentions*  
*Actions & Consequences*

# Reactive Speech Synthesis

C2H



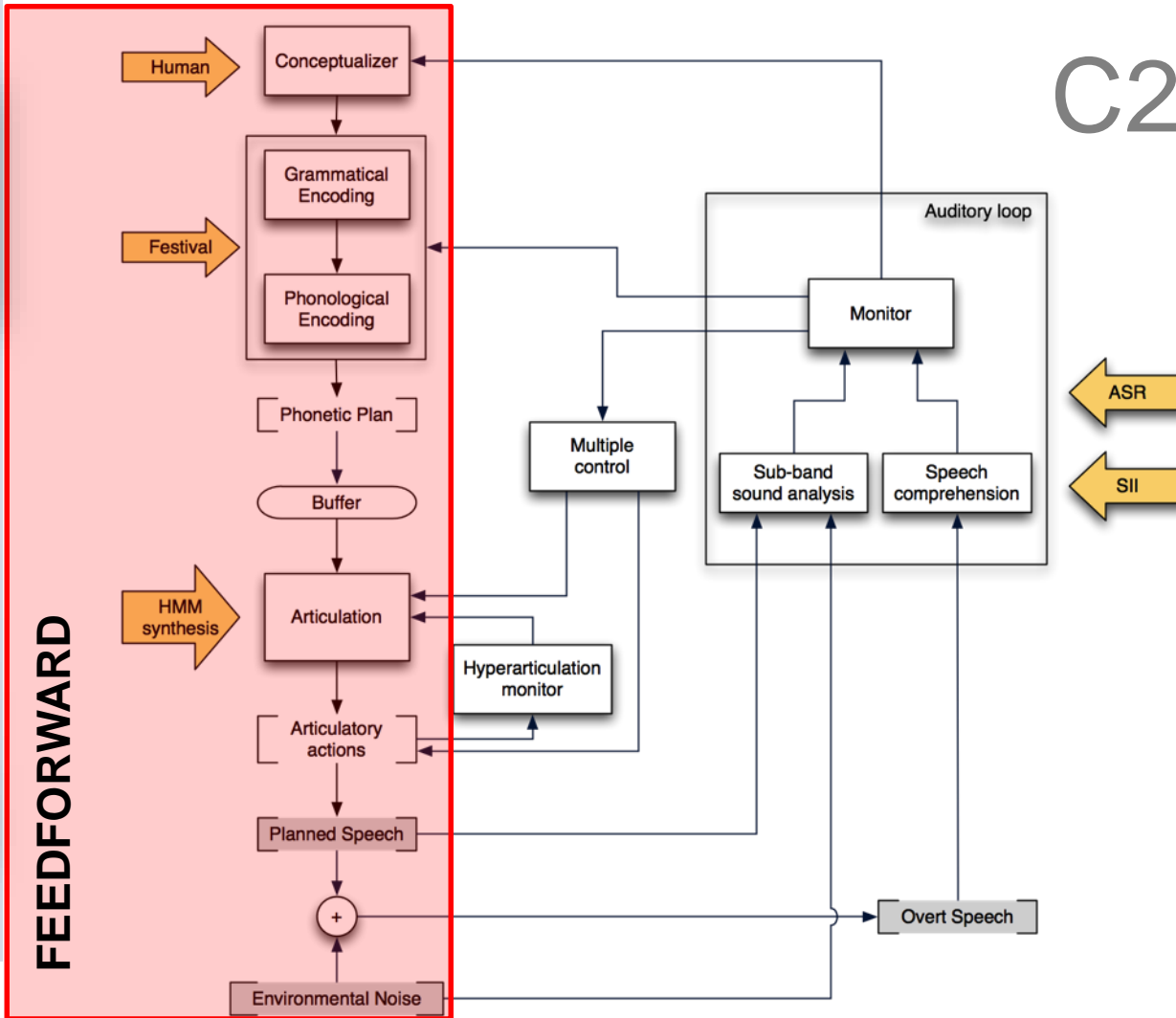
Mauro Nicolao



SCALE

# Reactive Speech Synthesis

C2H



Mauro Nicolao

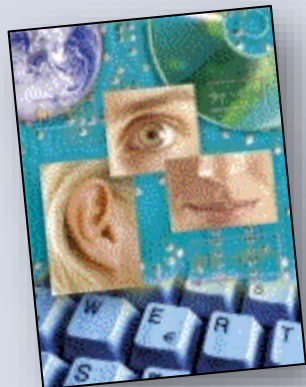


SCALE

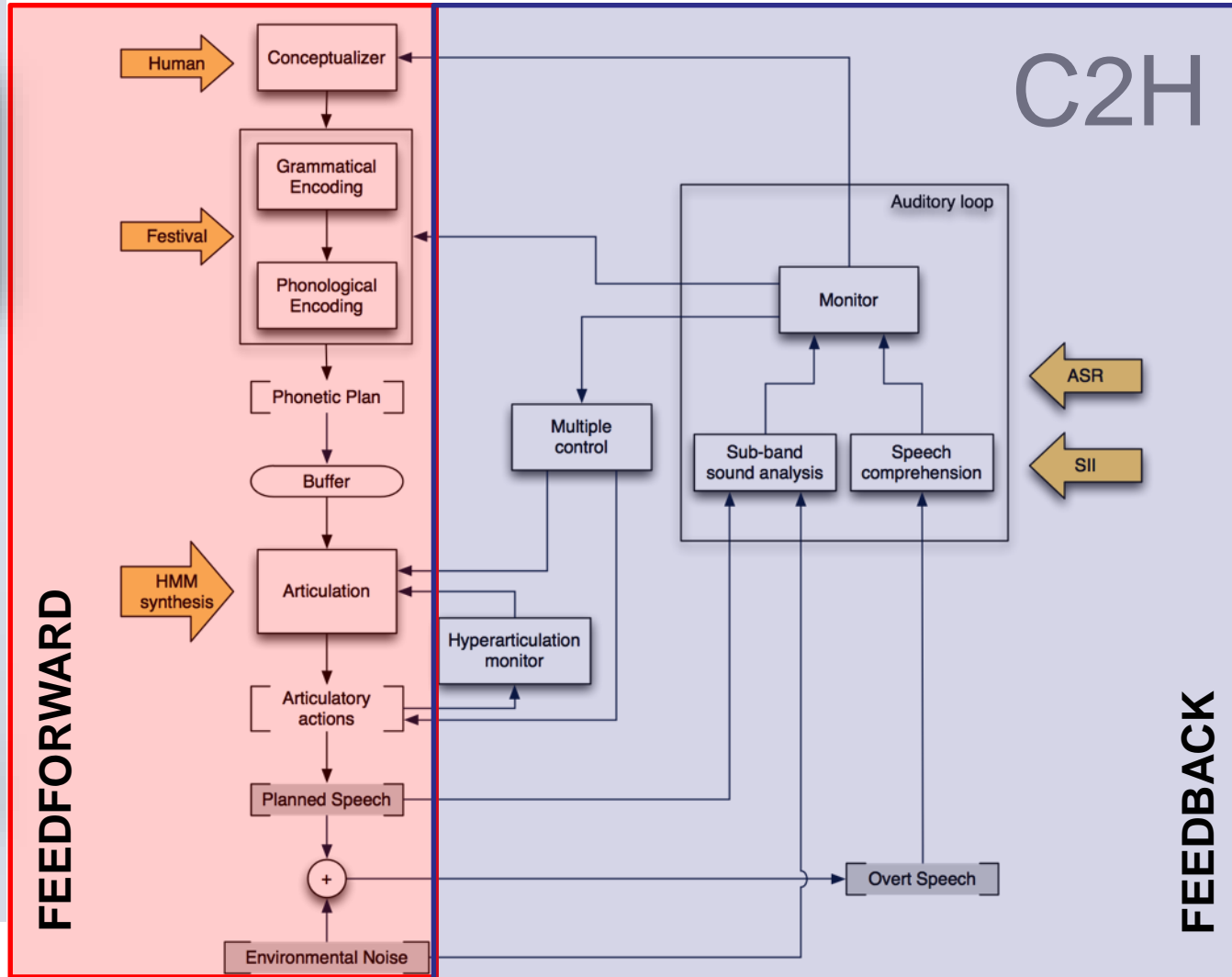
# Reactive Speech Synthesis



Mauro Nicolao



SCALE



# Reactive Speech Synthesis

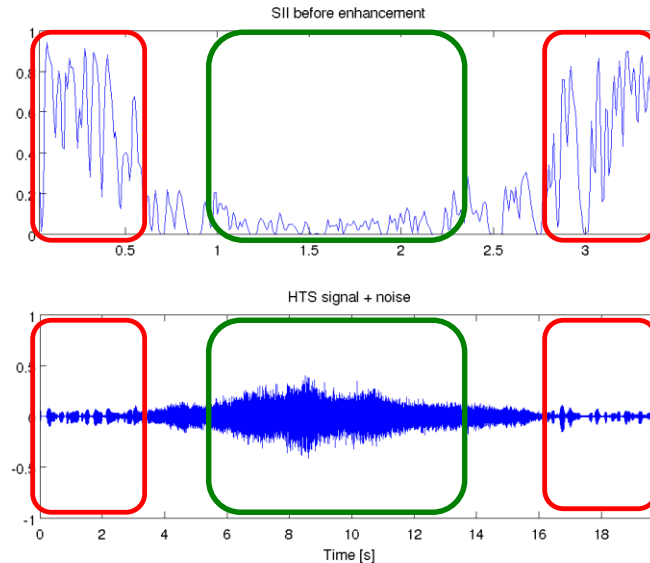


Mauro  
Nicolao

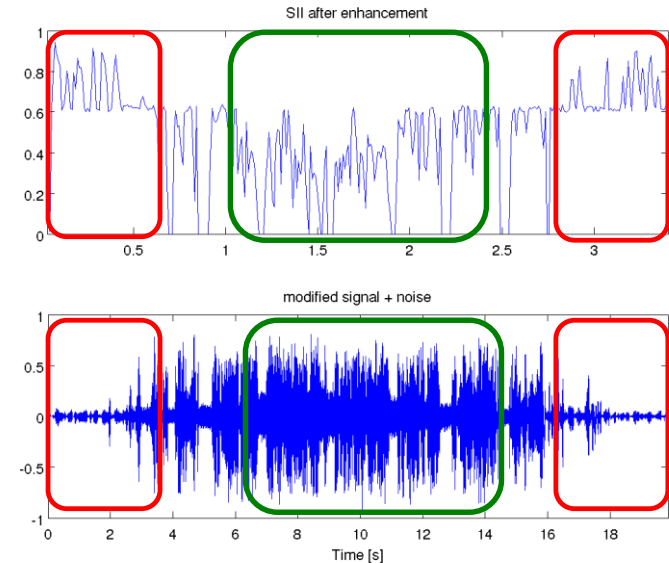


SCALE

## Traditional TTS



## Reactive TTS



Automatic compensation for disturbance

Moore, R. K., & Nicolao, M. (2011). Reactive speech synthesis: actively managing phonetic contrast along an H&H continuum, *17th International Congress of Phonetics Sciences (ICPhS)*. Hong Kong.



The  
University  
Of  
Sheffield.



# Reactive Speech Synthesis

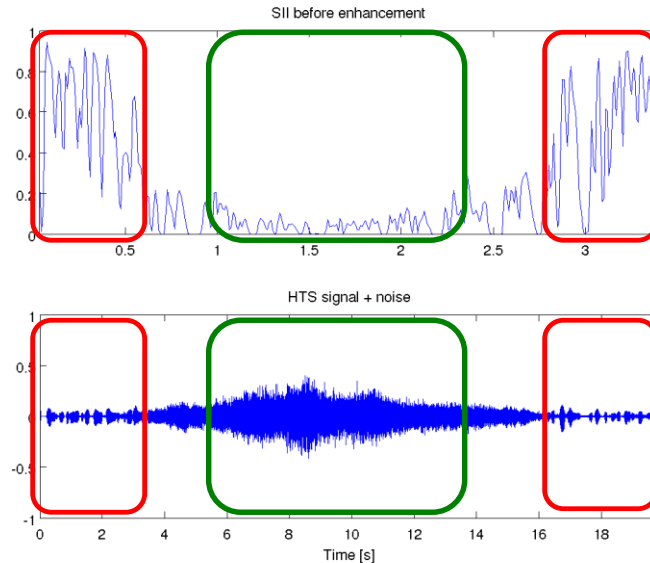


Mauro  
Nicolao

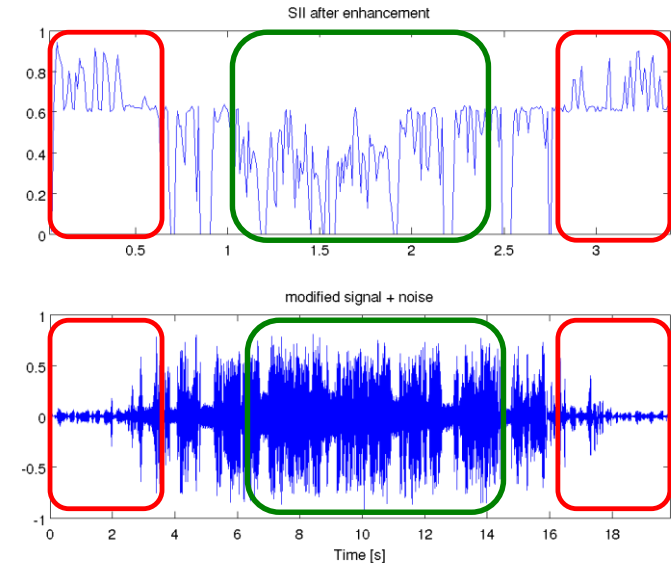


SCALE

## Traditional TTS



## Reactive TTS



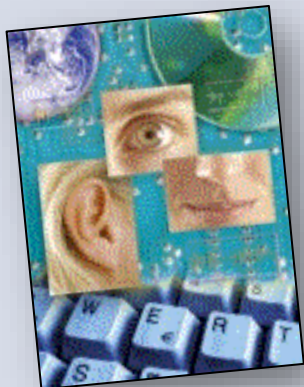
Automatic compensation for disturbance

Moore, R. K., & Nicolao, M. (2011). Reactive speech synthesis: actively managing phonetic contrast along an H&H continuum, *17th International Congress of Phonetics Sciences (ICPhS)*. Hong Kong.

# Reactive Speech Synthesis

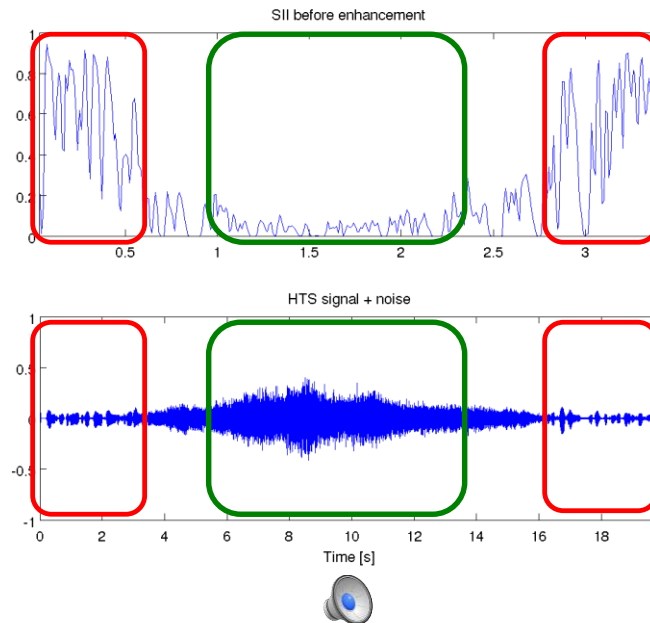


Mauro  
Nicolao

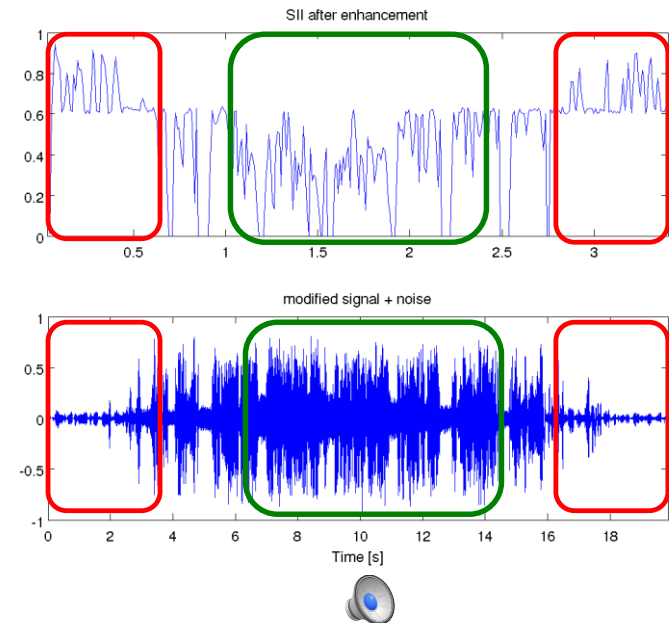


SCALE

## Traditional TTS



## Reactive TTS

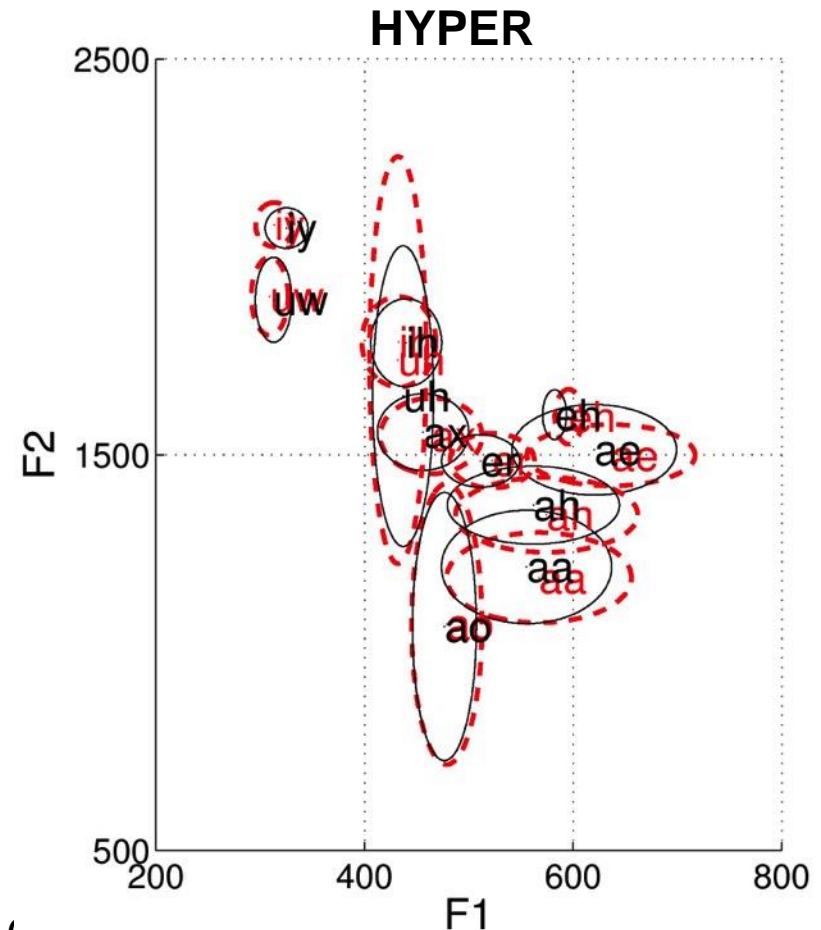
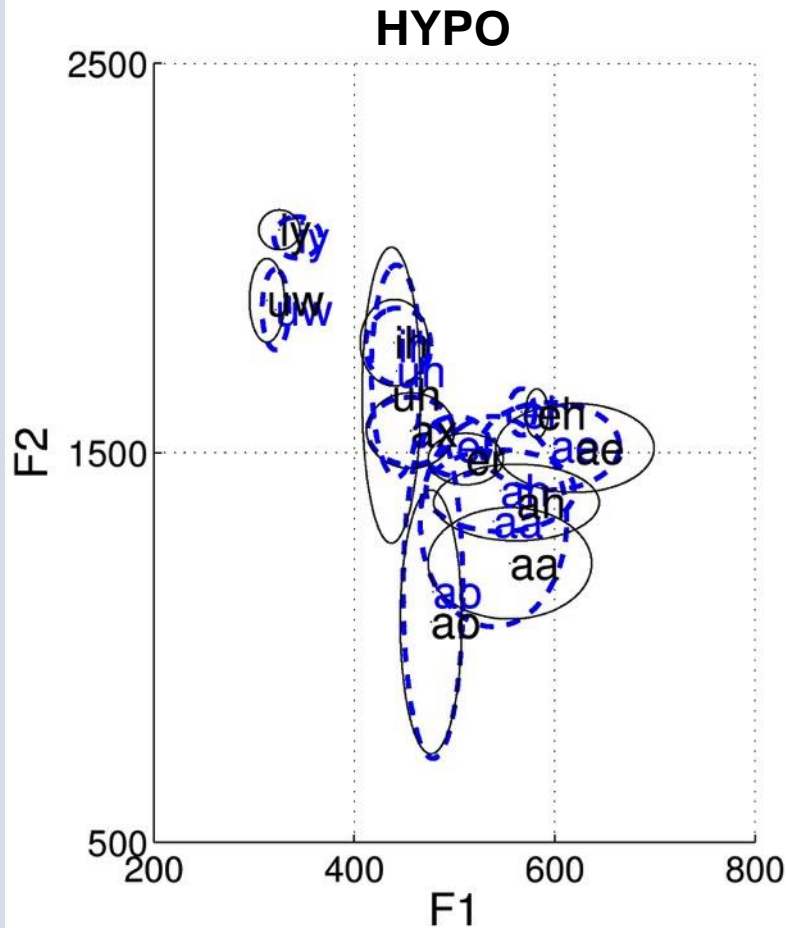


Automatic compensation for disturbance

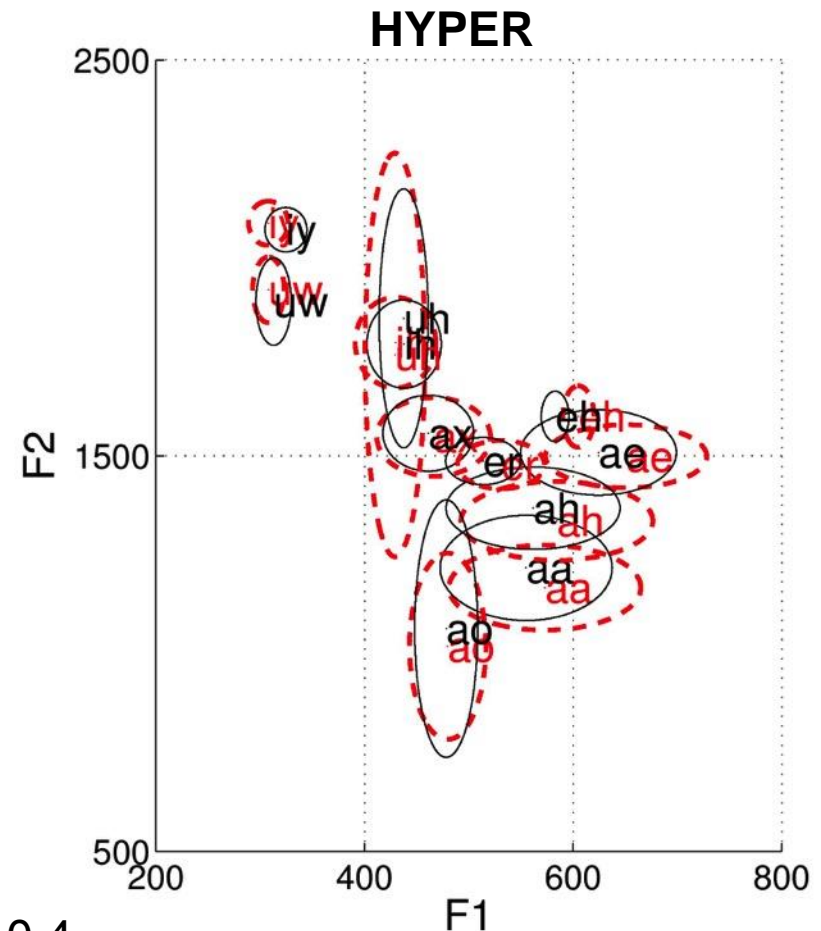
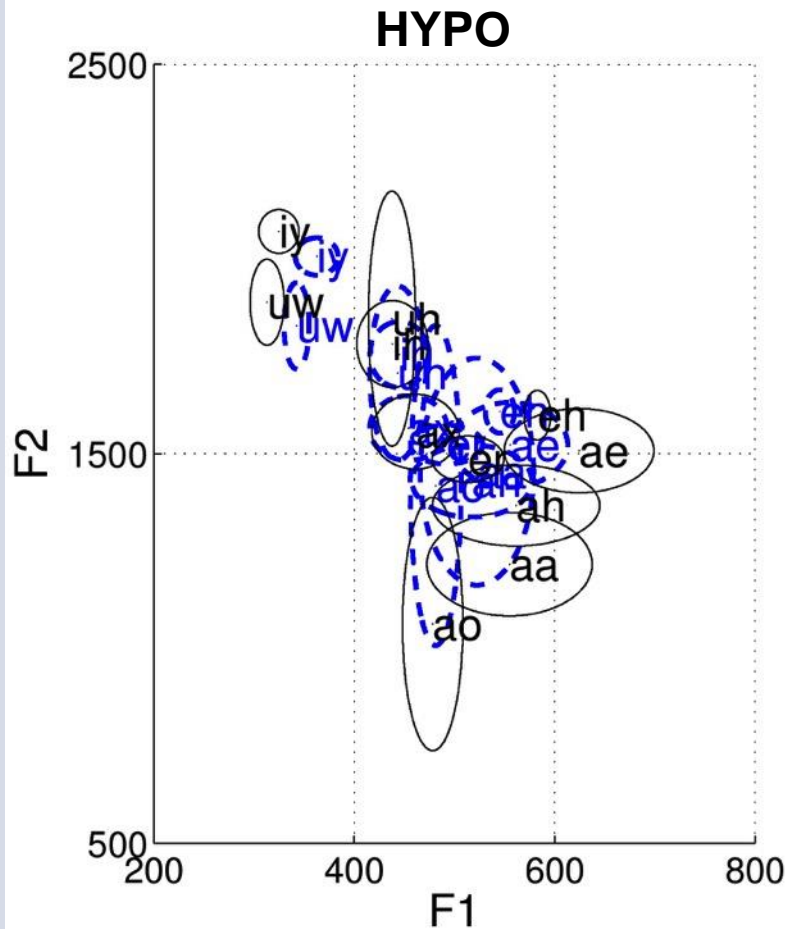
Moore, R. K., & Nicolao, M. (2011). Reactive speech synthesis: actively managing phonetic contrast along an H&H continuum, *17th International Congress of Phonetics Sciences (ICPhS)*. Hong Kong.



# Effect on Vowel Space

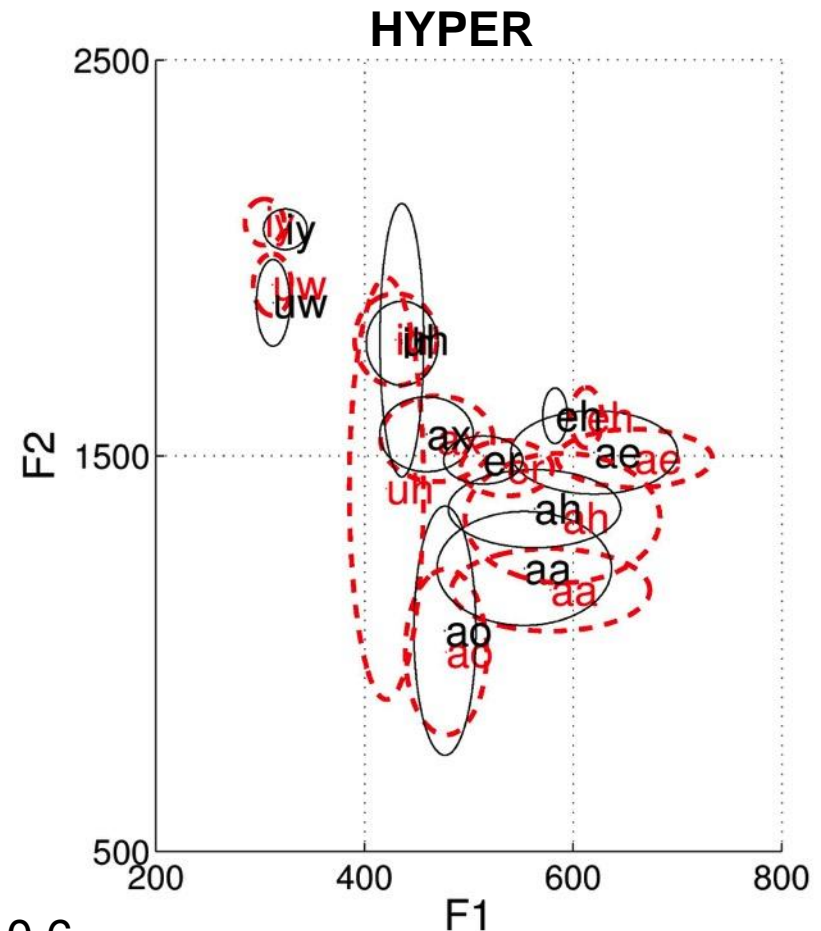
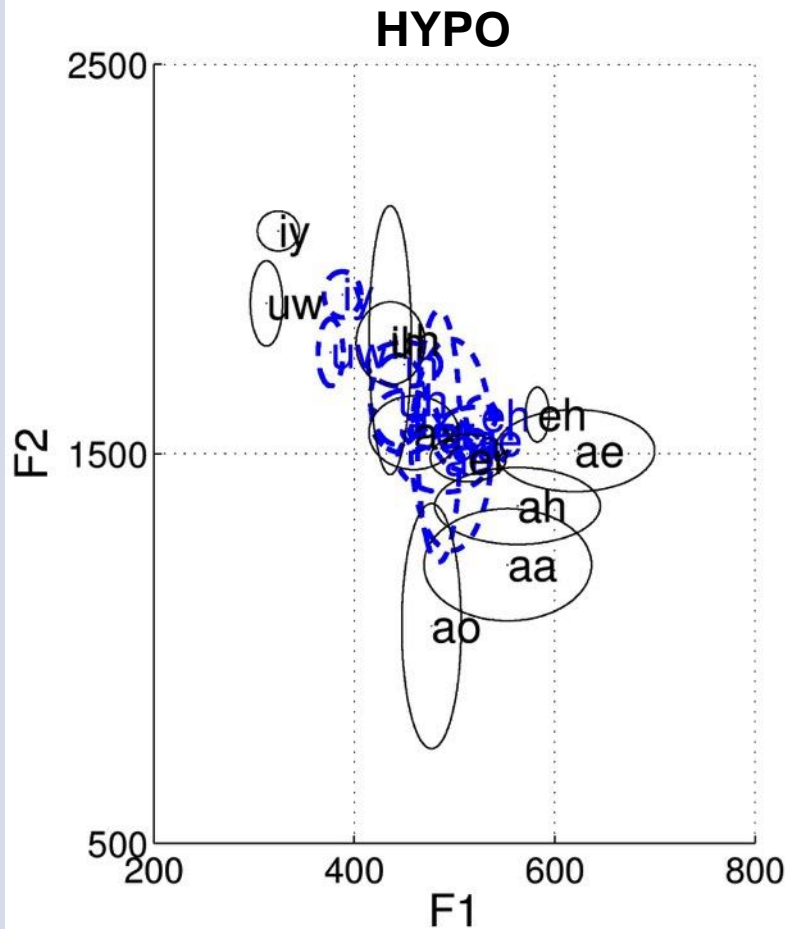

 $\alpha = ($

# Effect on Vowel Space

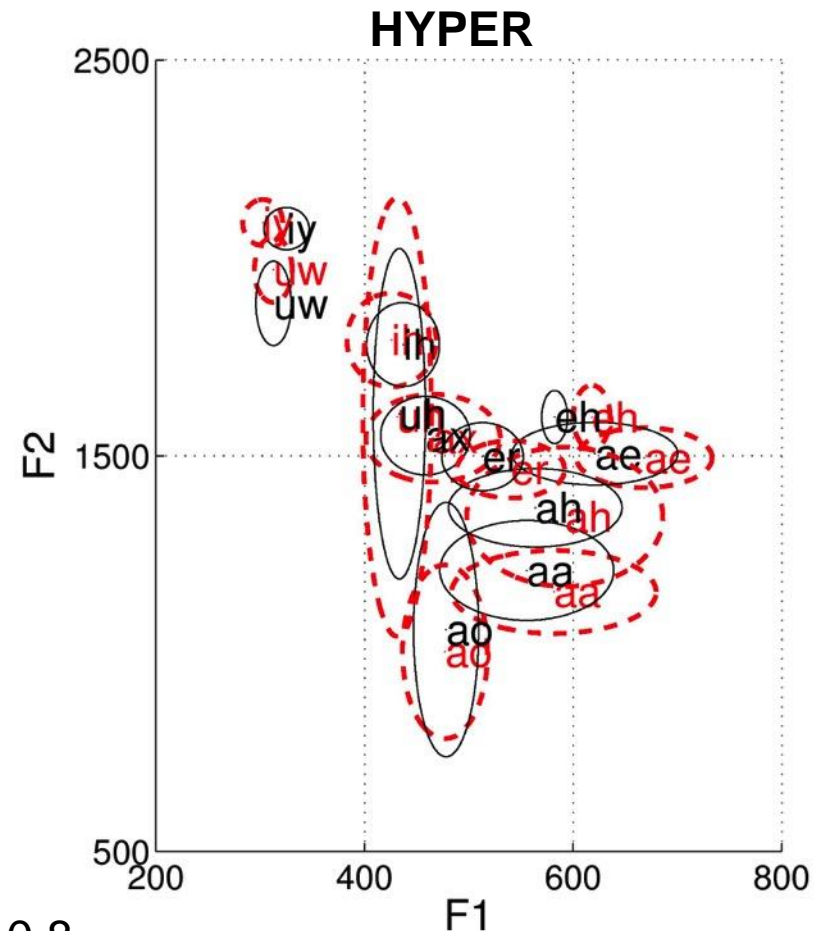
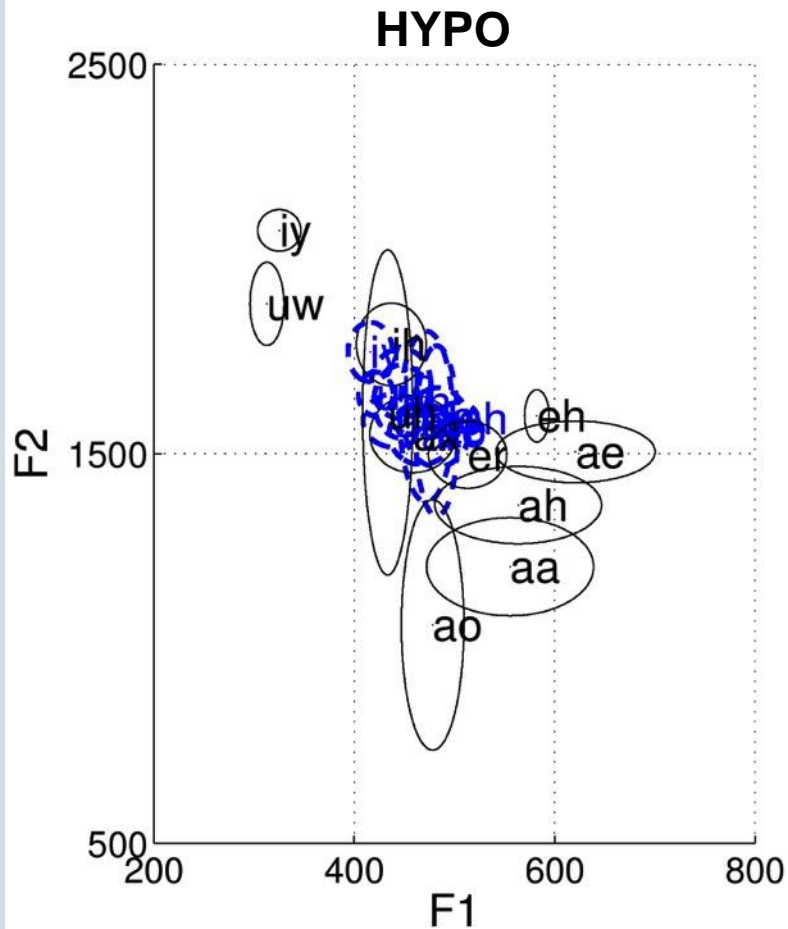


$\alpha = 0.4$

# Effect on Vowel Space

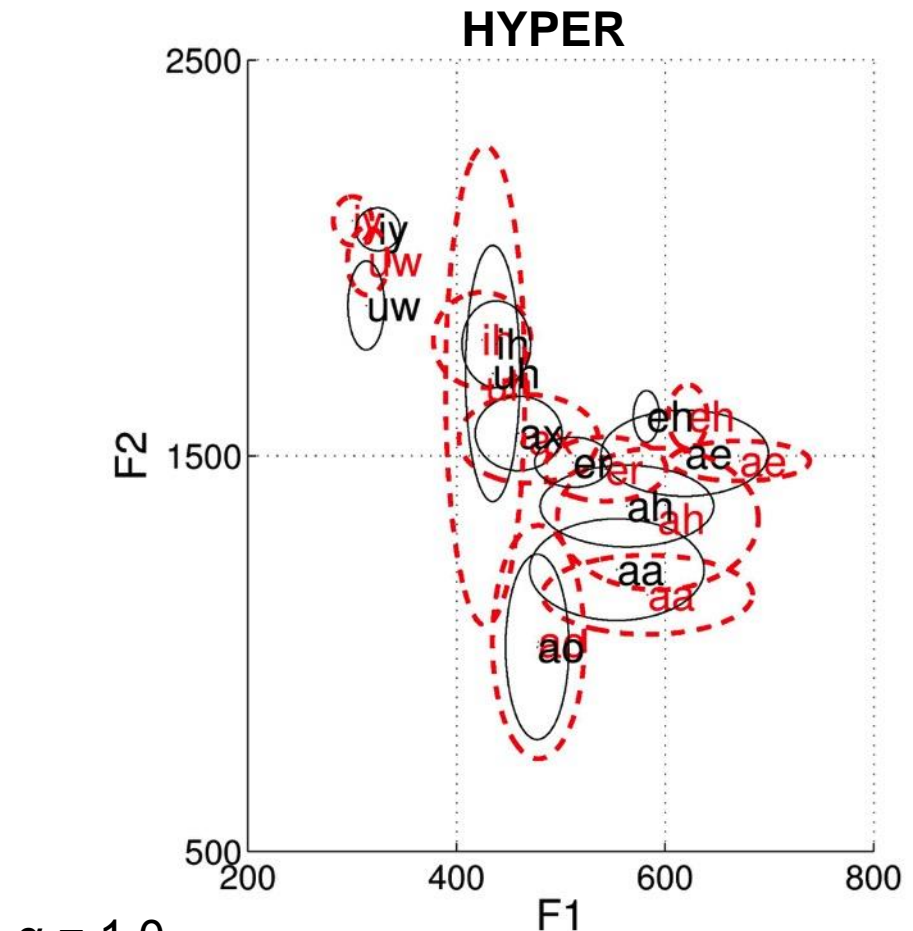
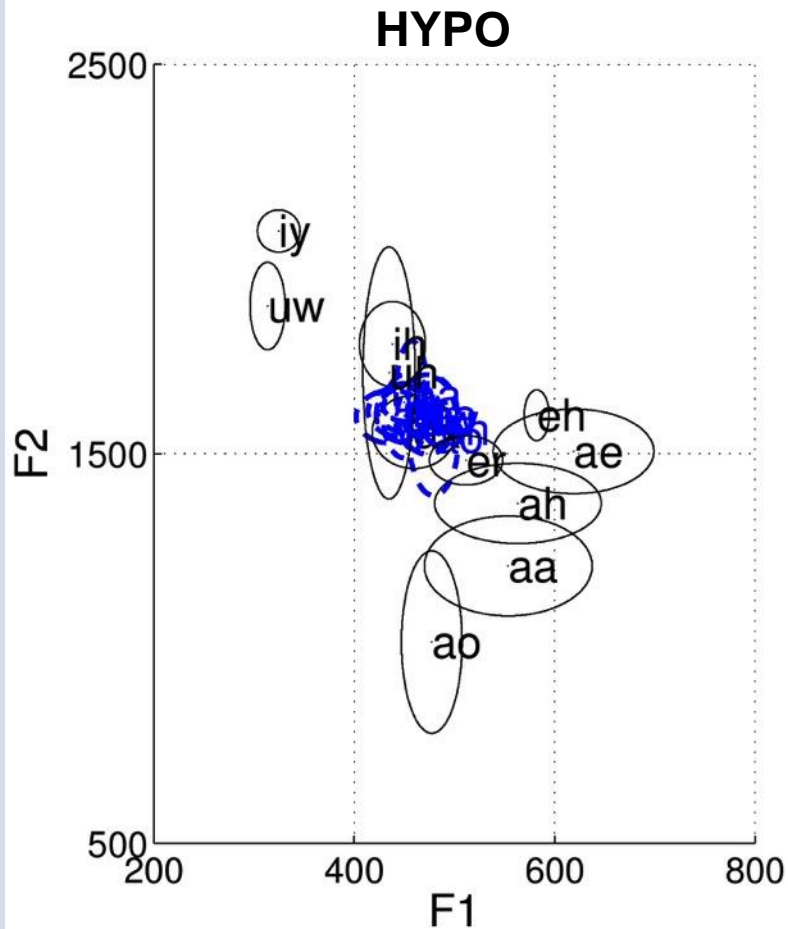

 $\alpha = 0.6$

# Effect on Vowel Space



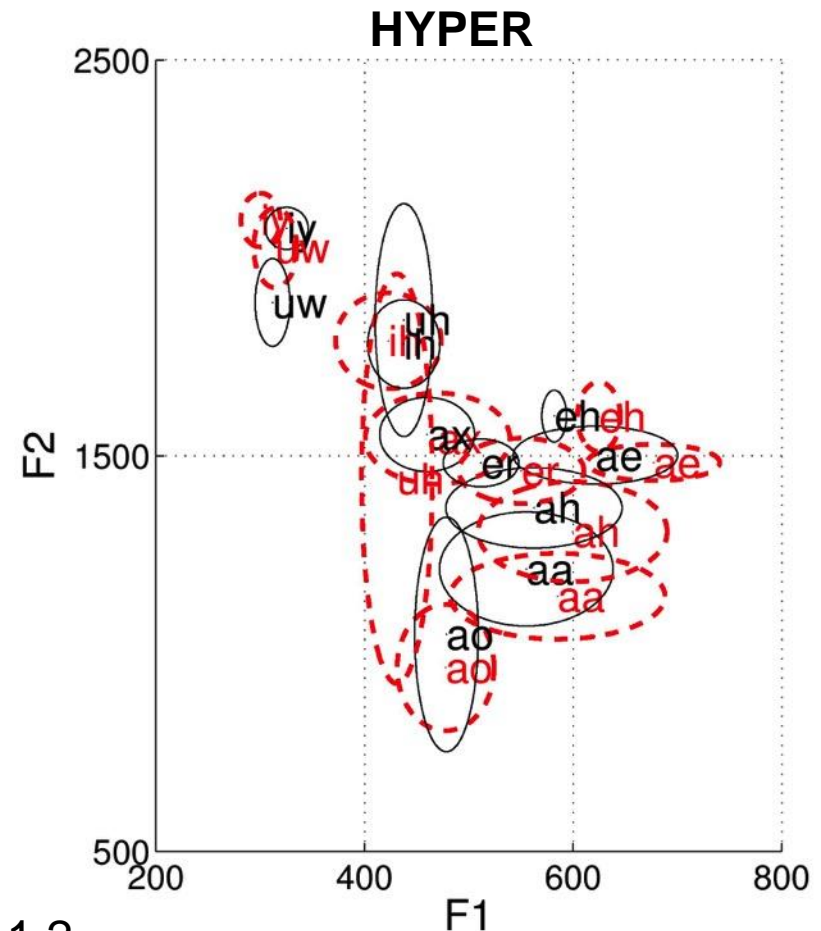
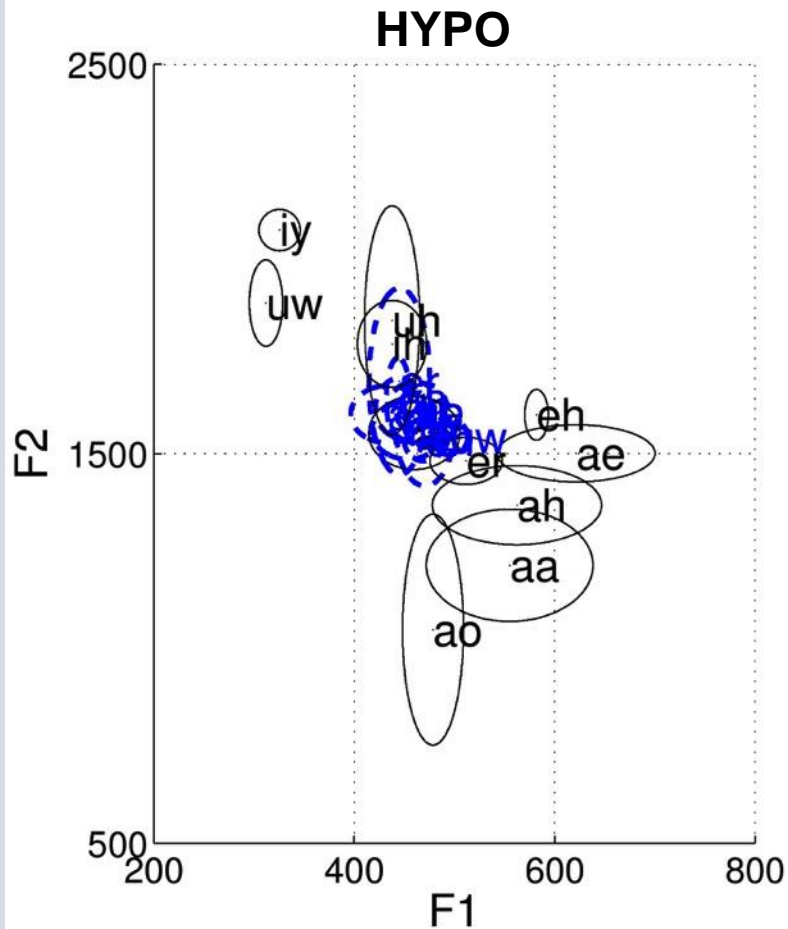
$\alpha = 0.8$

# Effect on Vowel Space

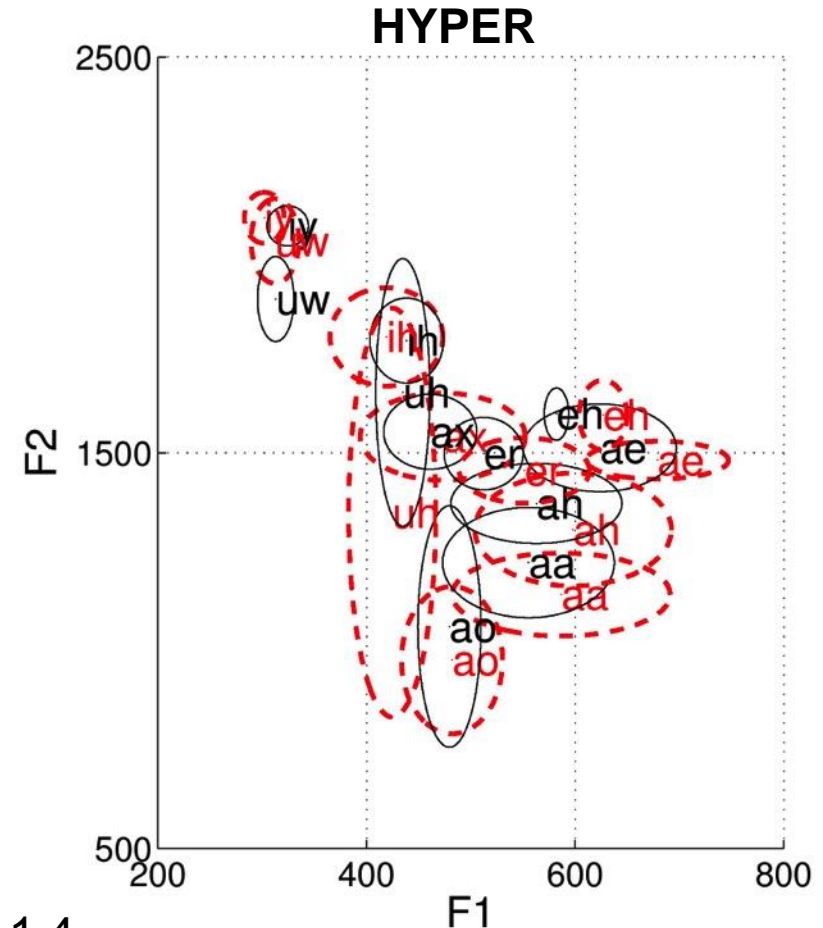
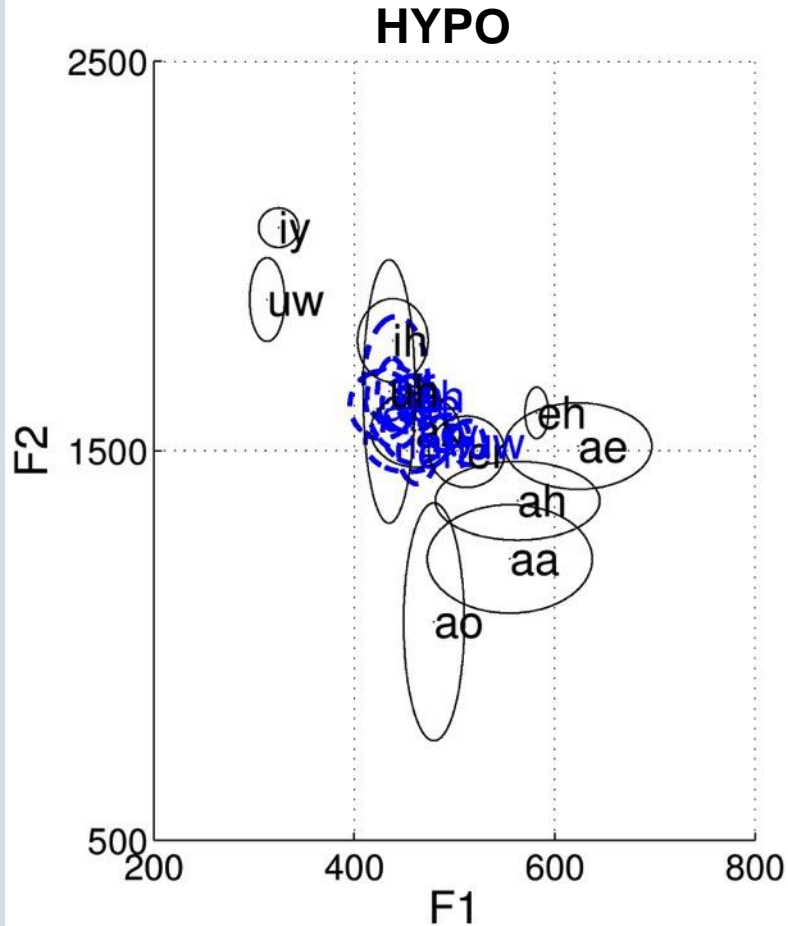


$\alpha = 1.0$

# Effect on Vowel Space

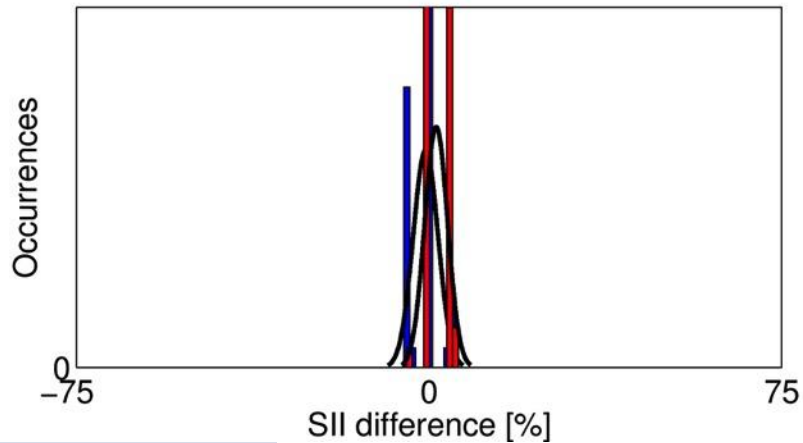

 $\alpha = 1.2$

# Effect on Vowel Space

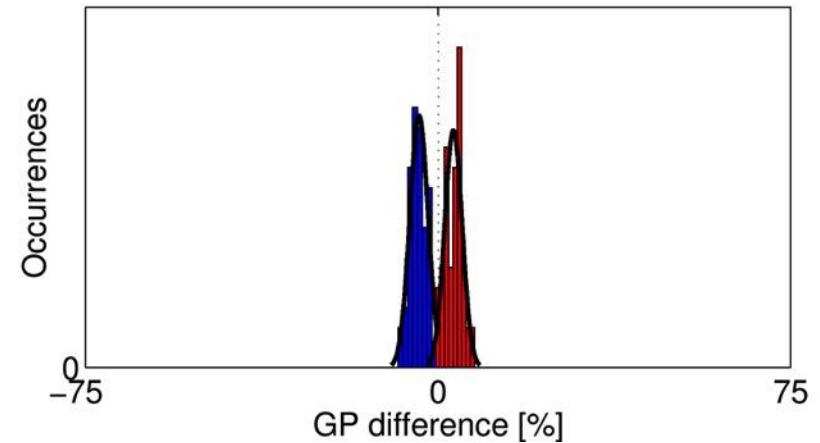

 $\alpha = 1.4$

# Effect on Intelligibility

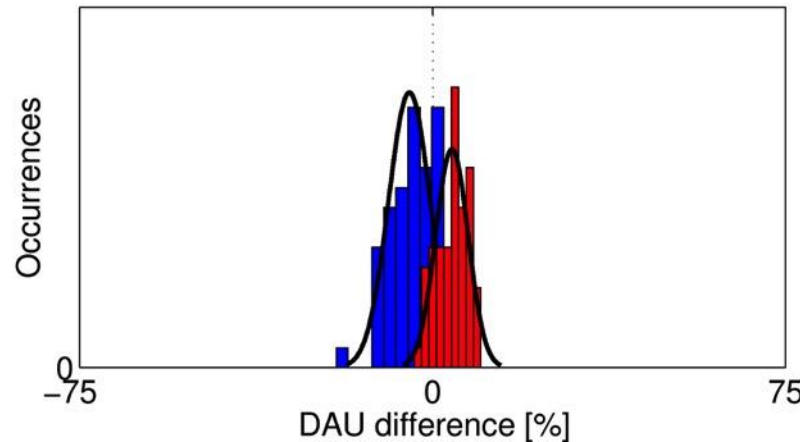
Mean<sub>HYO-STD</sub>: -0.72% Mean<sub>HYP-STD</sub>: +1.52%



Mean<sub>HYO-STD</sub>: -4.05% Mean<sub>HYP-STD</sub>: +3.27%



Mean<sub>HYO-STD</sub>: -4.91% Mean<sub>HYP-STD</sub>: +4.12%

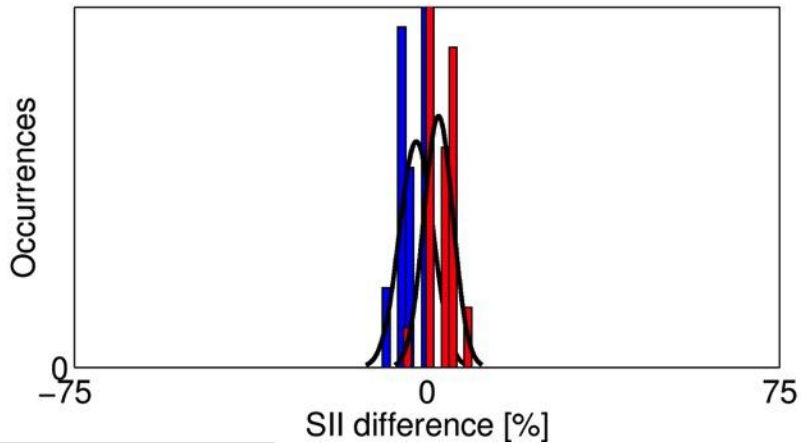


$\alpha = 0.2$

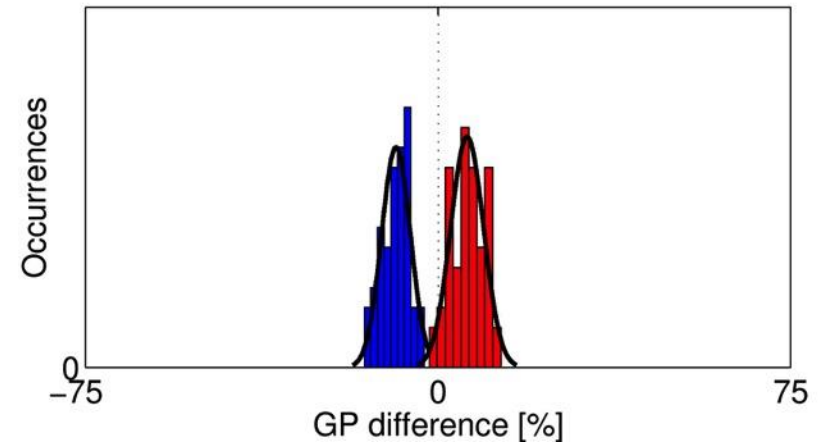


# Effect on Intelligibility

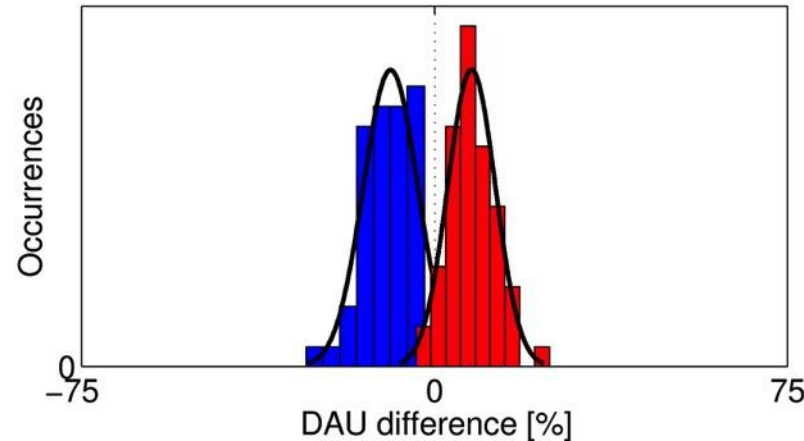
Mean<sub>HYO-STD</sub>: -2.27% Mean<sub>HYP-STD</sub>: +2.43%



Mean<sub>HYO-STD</sub>: -8.91% Mean<sub>HYP-STD</sub>: +6.18%



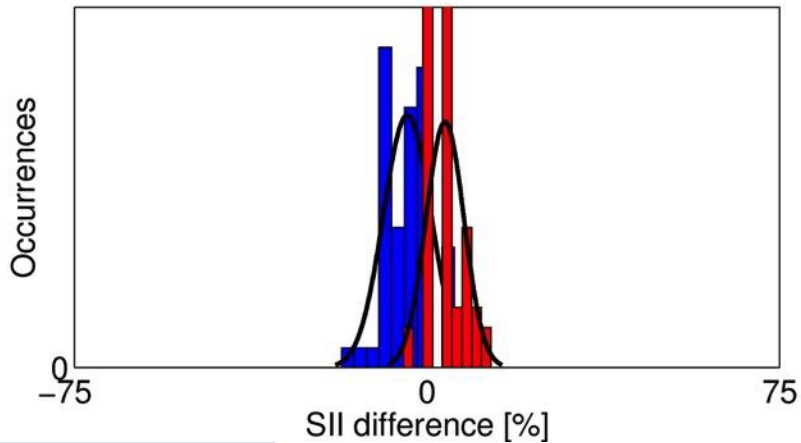
Mean<sub>HYO-STD</sub>: -9.33% Mean<sub>HYP-STD</sub>: +7.86%



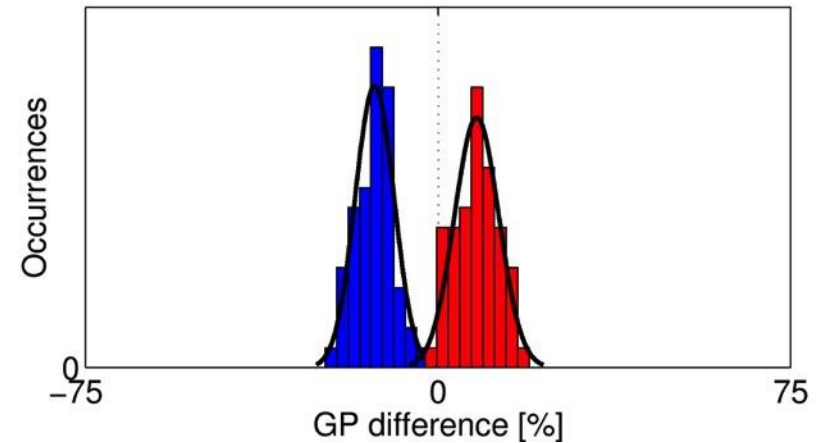
$\alpha = 0.4$

# Effect on Intelligibility

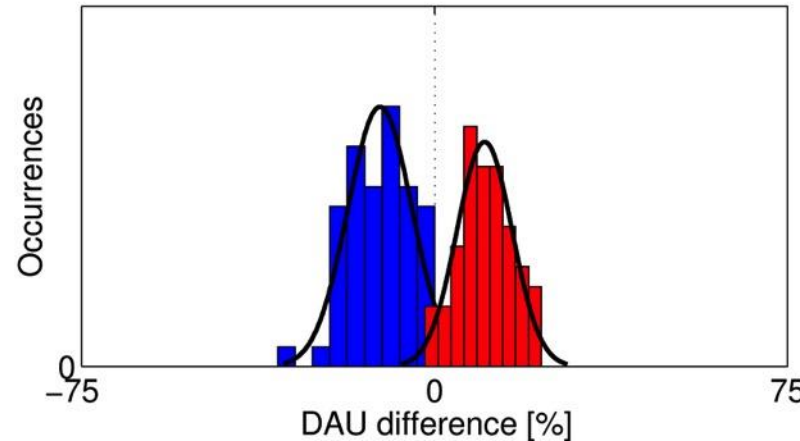
Mean<sub>HYO-STD</sub>: -4.18% Mean<sub>HYP-STD</sub>: +3.83%



Mean<sub>HYO-STD</sub>: -13.48% Mean<sub>HYP-STD</sub>: +8.22%



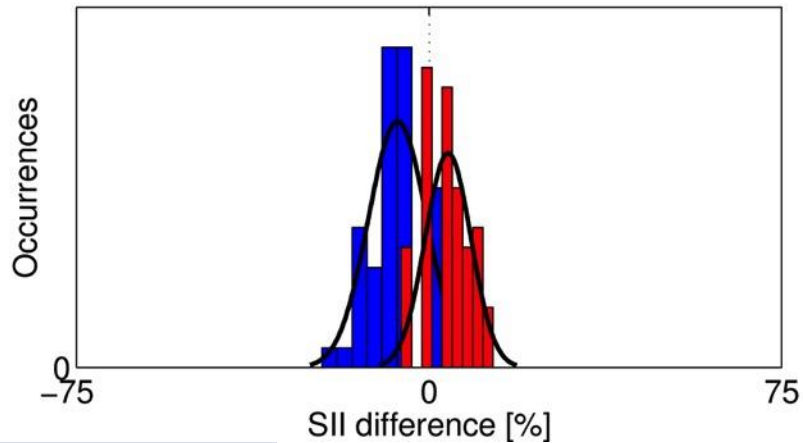
Mean<sub>HYO-STD</sub>: -11.58% Mean<sub>HYP-STD</sub>: +10.61%



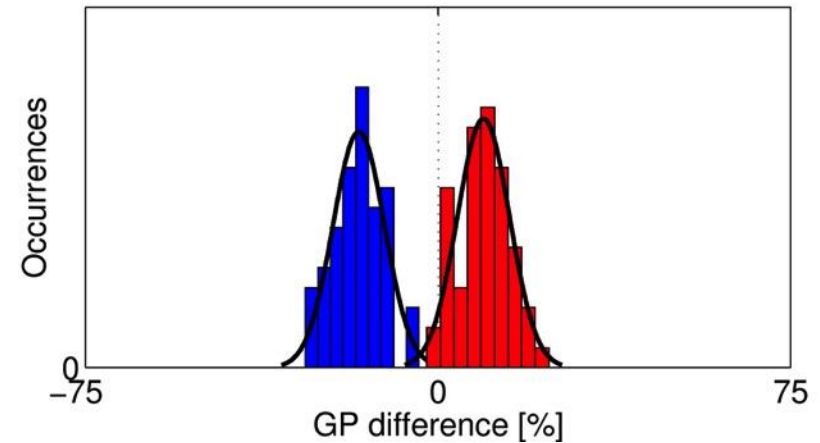
$\alpha = 0.6$

# Effect on Intelligibility

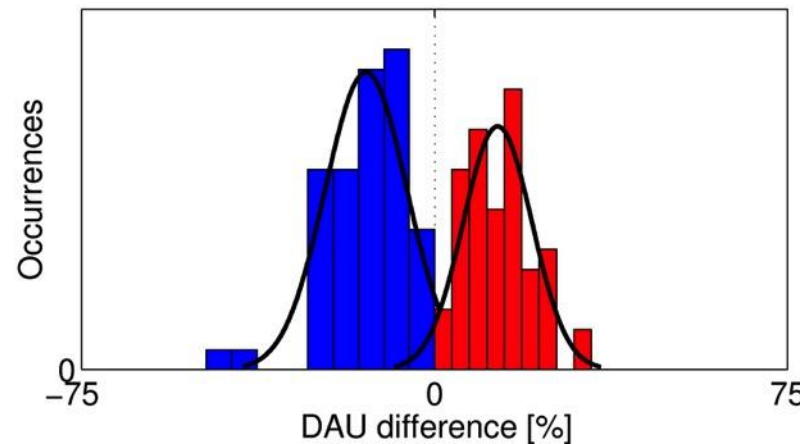
Mean<sub>HYO-STD</sub>: -6.68% Mean<sub>HYP-STD</sub>: +4.06%



Mean<sub>HYO-STD</sub>: -16.84% Mean<sub>HYP-STD</sub>: +9.67%



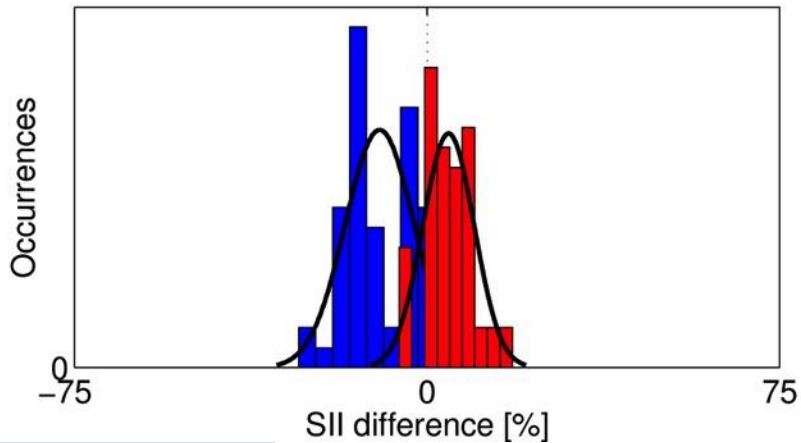
Mean<sub>HYO-STD</sub>: -14.53% Mean<sub>HYP-STD</sub>: +13.40%



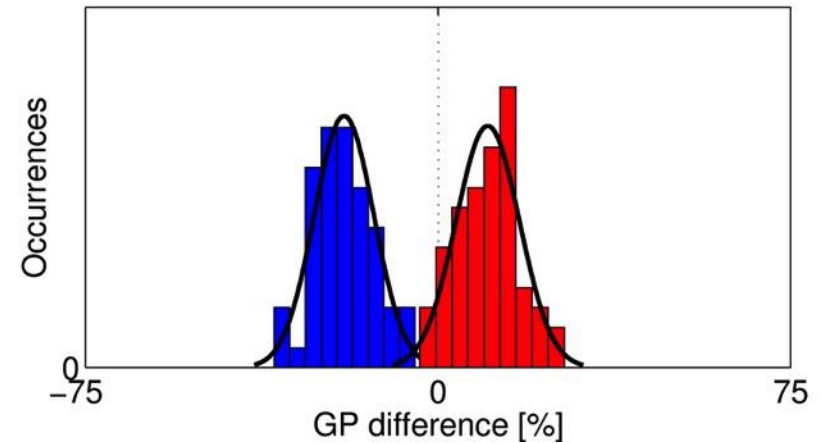
$\alpha = 0.8$

# Effect on Intelligibility

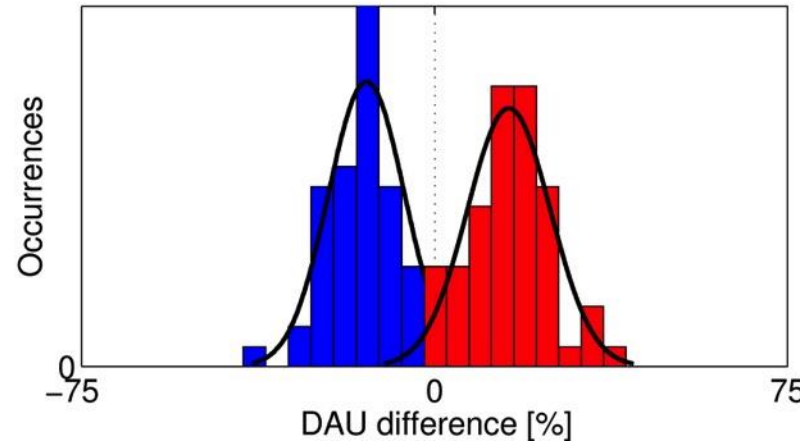
Mean<sub>HYO-STD</sub>: -10.07% Mean<sub>HYP-STD</sub>: +4.58%



Mean<sub>HYO-STD</sub>: -19.97% Mean<sub>HYP-STD</sub>: +10.56%



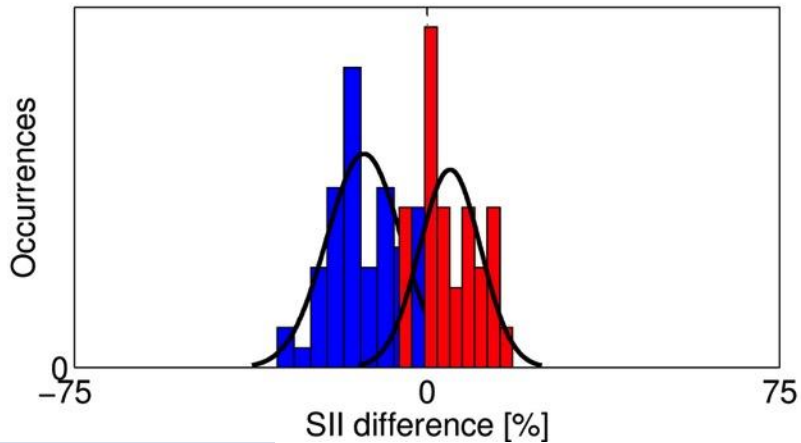
Mean<sub>HYO-STD</sub>: -14.44% Mean<sub>HYP-STD</sub>: +15.79%



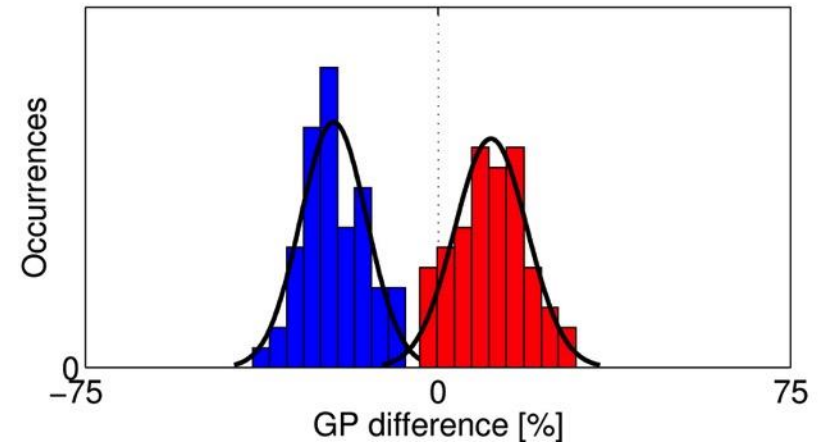
$\alpha = 1.0$

# Effect on Intelligibility

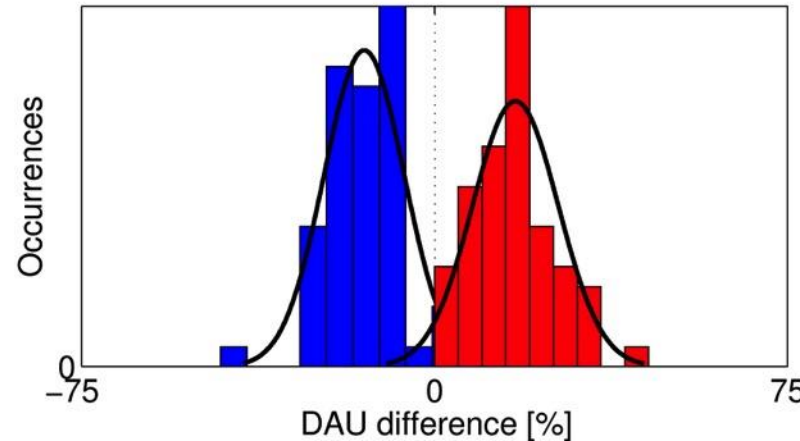
Mean<sub>HYO-STD</sub>: -13.38% Mean<sub>HYP-STD</sub>: +4.91%



Mean<sub>HYO-STD</sub>: -22.20% Mean<sub>HYP-STD</sub>: +11.26%



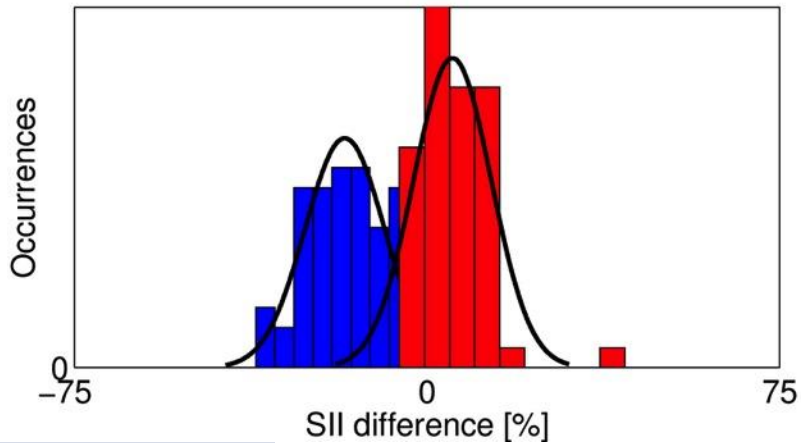
Mean<sub>HYO-STD</sub>: -14.94% Mean<sub>HYP-STD</sub>: +17.18%



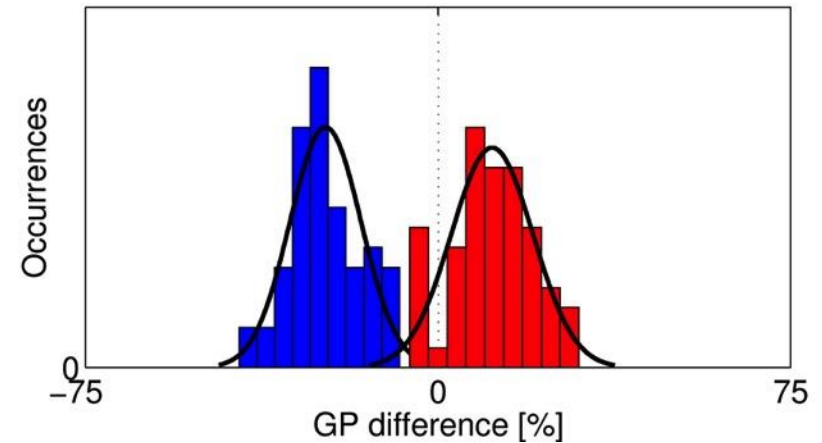
$\alpha = 1.2$

# Effect on Intelligibility

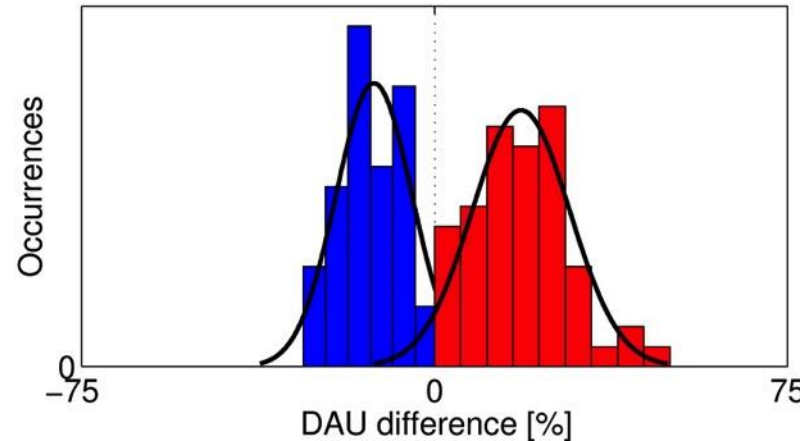
Mean<sub>HYO-STD</sub>: -17.48% Mean<sub>HYP-STD</sub>: +5.40%



Mean<sub>HYO-STD</sub>: -23.91% Mean<sub>HYP-STD</sub>: +11.50%












Mean<sub>HYO-STD</sub>: -12.85% Mean<sub>HYP-STD</sub>: +18.36%















$\alpha = 1.4$

# Reactive Speech Synthesis

Type of noise	HYPO	NORM	HYPER
Speech Shaped Noise (SNR = 1 dB)			
Competing Talker (SNR = -7 dB)			
Clean			

*“The box was thrown beside  
the parked truck”*

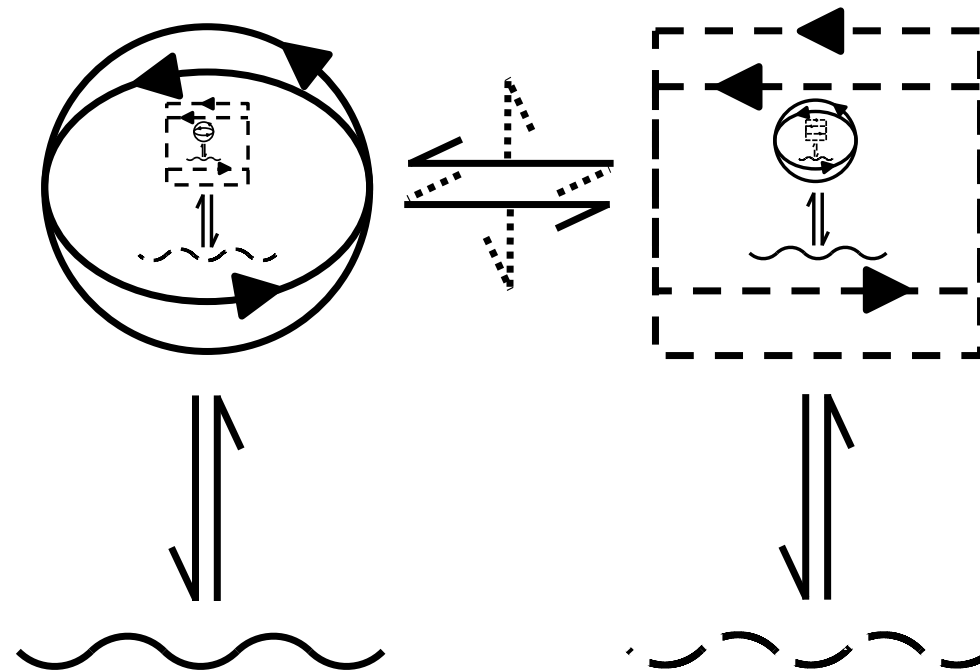
# Reactive Speech Synthesis

Type of noise	HYPO	NORM	HYPER
Car Noise (SNR = -4 dB)			
Babble Noise (SNR = -4 dB)			
Competing Talkers (SNR = -4 dB)			
Clean			

Nicolao, M., Tesser, F., & Moore, R. K. (2013). A phonetic-contrast motivated adaptation to control the degree-of-articulation on Italian HMM-based synthetic voices. In *8th ISCA Speech Synthesis Workshop (SSW8)*. Barcelona, Spain.



# Concluding Remarks

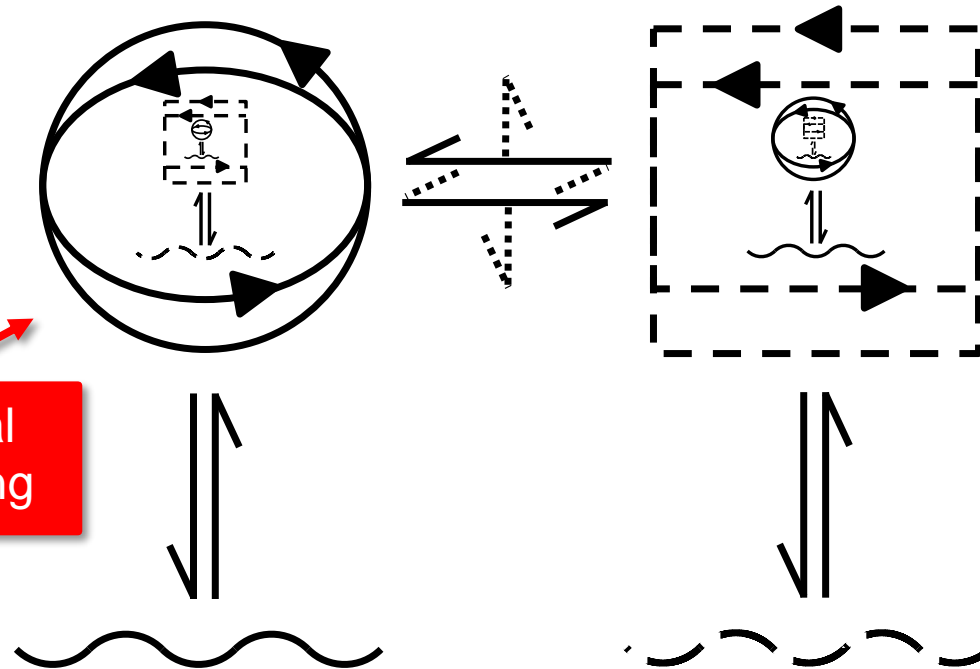


Moore, R. K. (2016). Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *Int. J. Advanced Robotic Systems*, 13(74).

# Concluding Remarks

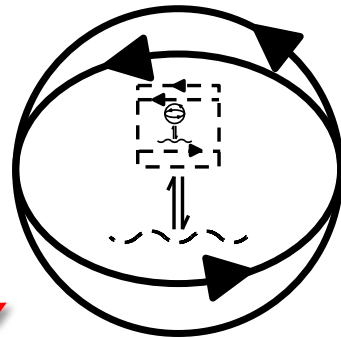


Ostensive inferential recursive mindreading

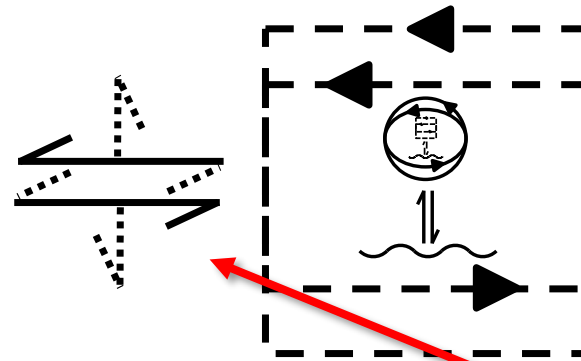


Moore, R. K. (2016). Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *Int. J. Advanced Robotic Systems*, 13(74).

# Concluding Remarks



Ostensive inferential recursive mindreading



Mutual declarative, interrogative, imperative coupling

Moore, R. K. (2016). Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *Int. J. Advanced Robotic Systems*, 13(74).



Thank You

<http://www.dcs.shef.ac.uk/~roger>

- The field of spoken language processing typically treats speech as classic stimulus-response behaviour, hence there is strong interest in using the latest machine learning techniques (such as Deep Neural Networks) to estimate the assumed non-linear transforms.
- However, in reality, speech is not a static process - rather it is a sophisticated joint behaviour resulting from actively managed dynamic coupling between speakers, listeners and their respective environmental contexts.
- Multiple layers of feedback control play a crucial role in maintaining the necessary communicative stability, and this means that there are significant dependencies that are overlooked in contemporary SLP approaches.
- This talk will address these issues in the wider context of intentional behaviour, and will give an insight into the implications for computational models.