Developing a computational model of soundchange

A developmental model of vowel systems using self-organisation

Raphael Winkelmann
raphael@phonetik.uni-muenchen.de

Institute of Phonetics and Speech Processing , Ludwig-Maximilians-Universität München

for the seminar @ V.I.U.:
The relationships between speech production and speech perception

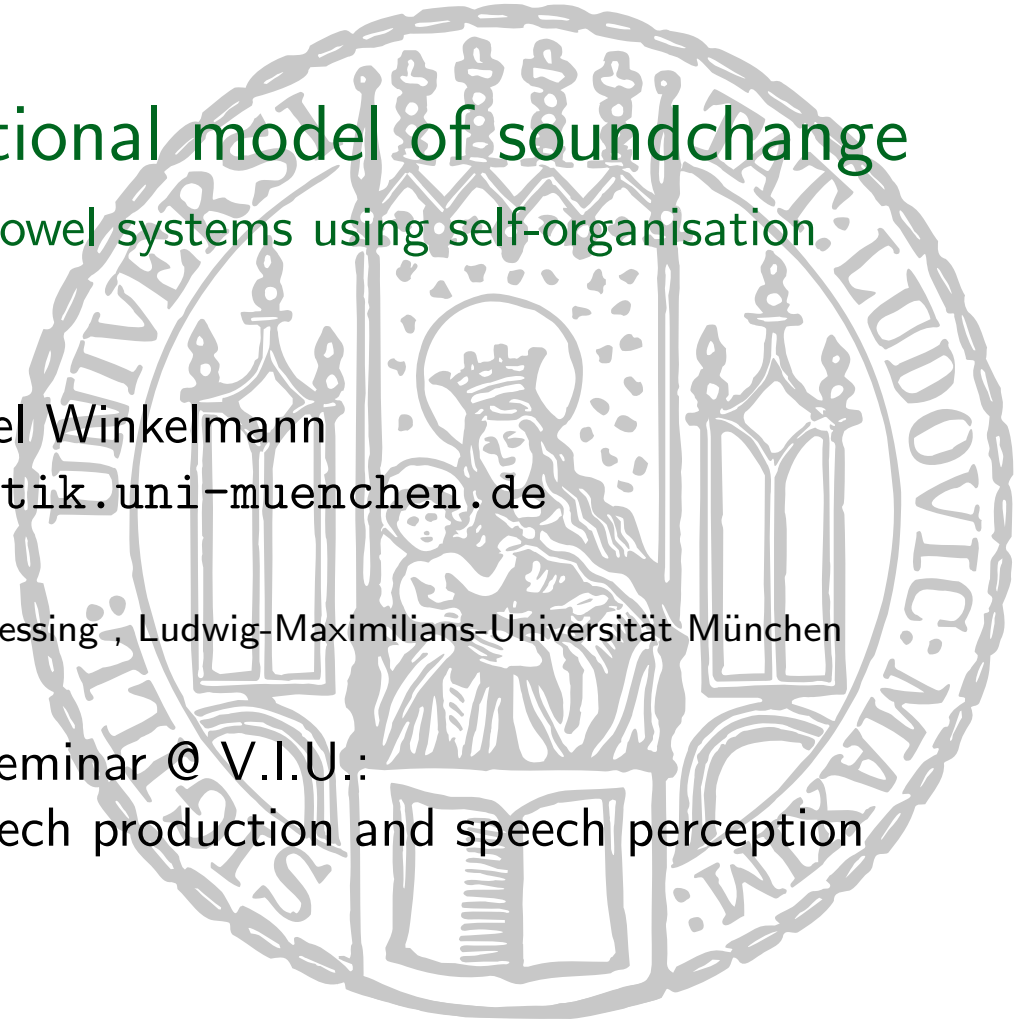


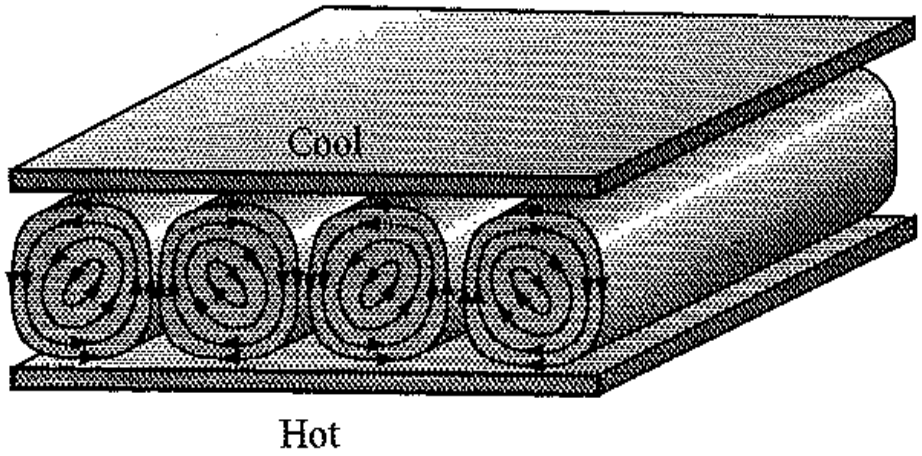
table of contents

The principle of self-organisation

- Definition:
  - Self-organization is the process where a structure or pattern appears in a system without a central authority or external element imposing it through planning.[1]

[1] http://en.wikipedia.org/wiki/Self-organization

The principle of self-organisation (Example: Rayleigh-Bénard-Convection[2]):



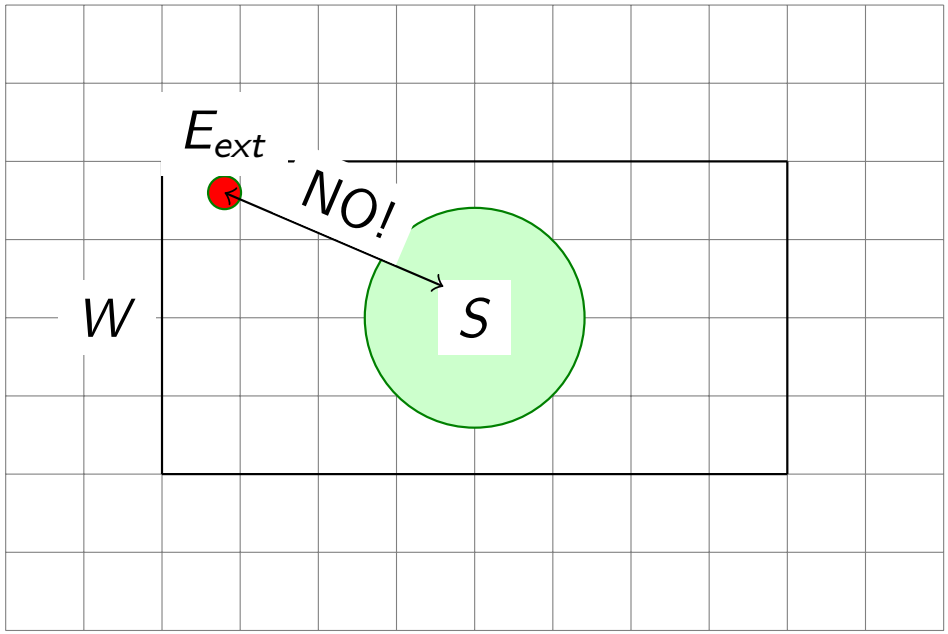
If a thin layer of liquid is heated on a stove, then given a certain minimum temperature difference between the top and the bottom of the liguid, there is self-organization of convection currents in parallel stripes

[2] from Oudeyer 2006

System

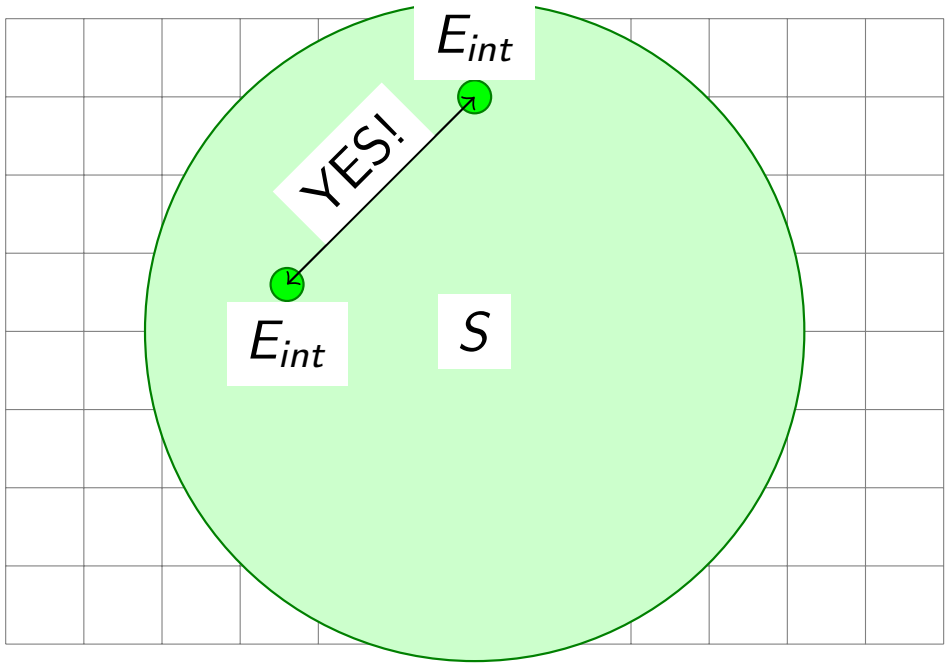
Prerequisites:

- A system with entities that are able to/must interact.
- System is autonomous → no outside interaction



$W = world$
$S = system$
$E = entity_{external}$

Agents as entities in the system

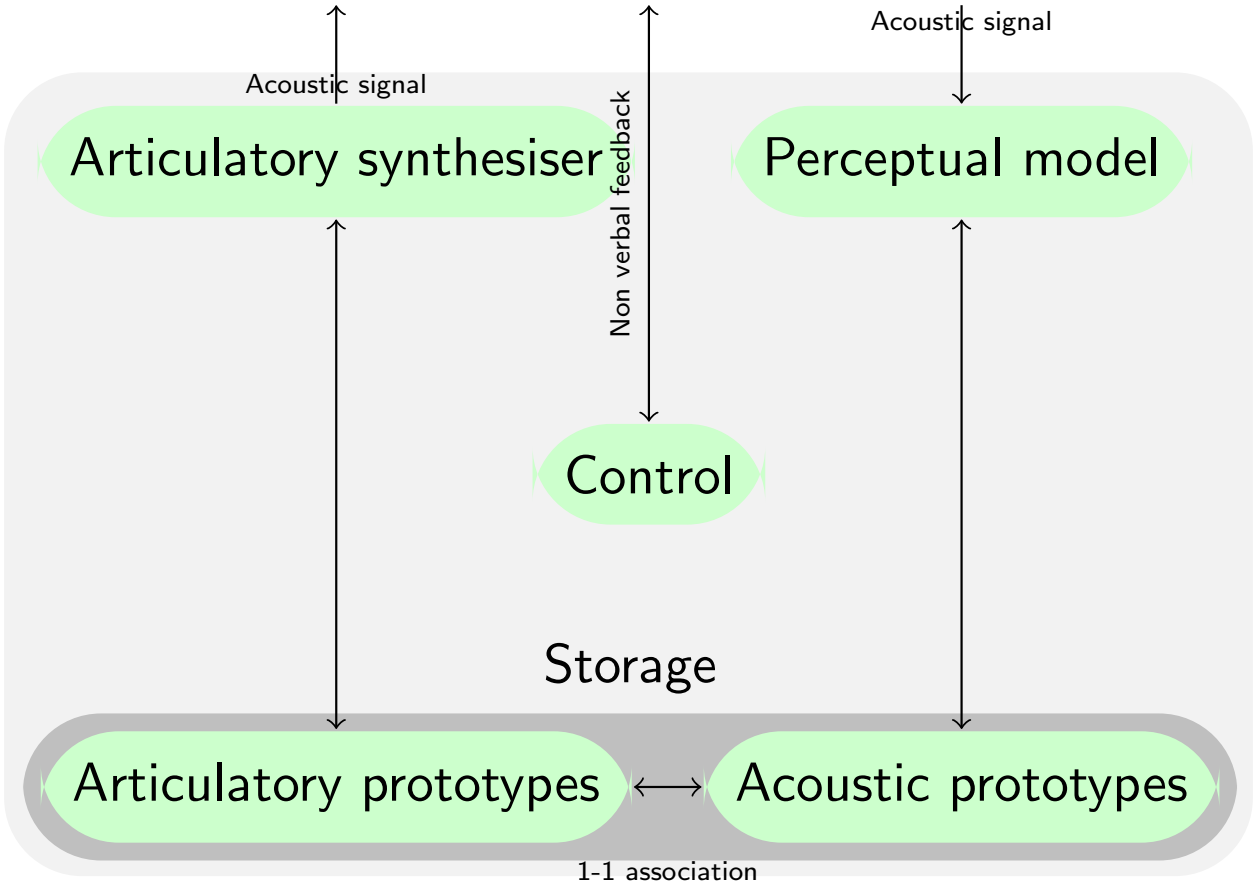


$S = system$
$E = entity_{internal}$

Agents as entities in the system

- Agents represent an abstract entity with partial humanoid capabilities
- Properties:
  - Vowel system (articulatory prototypes)
  - Articulatory synthesiser
  - A human perception model
- All the agents within a system represents the population P

Agent architecture[3]



[3] Adapted from Bart de Boer 2000

Storage: Data Points[4]

Vowel	p	h	r	F_1(Hz)	F_2(Hz)	F_3(Hz)	F_4(Hz)
[a]	0	0	0	708	1517	2427	3678
[œ]	0	0	1	670	1400	2300	3500
[ɐ]	0.5	0	0	742	1266	2330	3457
[ɐ̟]	0.5	0	1	658	1220	2103	3200
[ɑ]	1	0	0	703	1074	2356	3486
[ɒ]	1	0	1	656	1020	2312	3411
[e]	0	0.5	0	395	2027	2552	3438
[ø]	0	0.5	1	393	1684	2238	3254
[ə]	0.5	0.5	0	399	1438	2118	3197
[e]	0.5	0.5	1	400	1267	2005	2996
[ɣ]	1	0.5	0	430	1088	2142	3490
[o]	1	0.5	1	399	829	2143	3490
[i]	0	1	0	252	2202	3242	3938
[y]	0	1	1	250	1878	2323	3447
[ɨ]	0.5	1	0	264	1591	2259	3592
[ʉ]	0.5	1	1	276	1319	2082	3118
[ɯ]	1	1	0	305	1099	2220	3604
[u]	1	1	1	276	740	2177	3506

where p = tongue position; h = tongue height; r = lip rounding (all elements of the interval between 0 and 1). These represent the the three major vowel parameters.

[4] Data from Vallée 1994

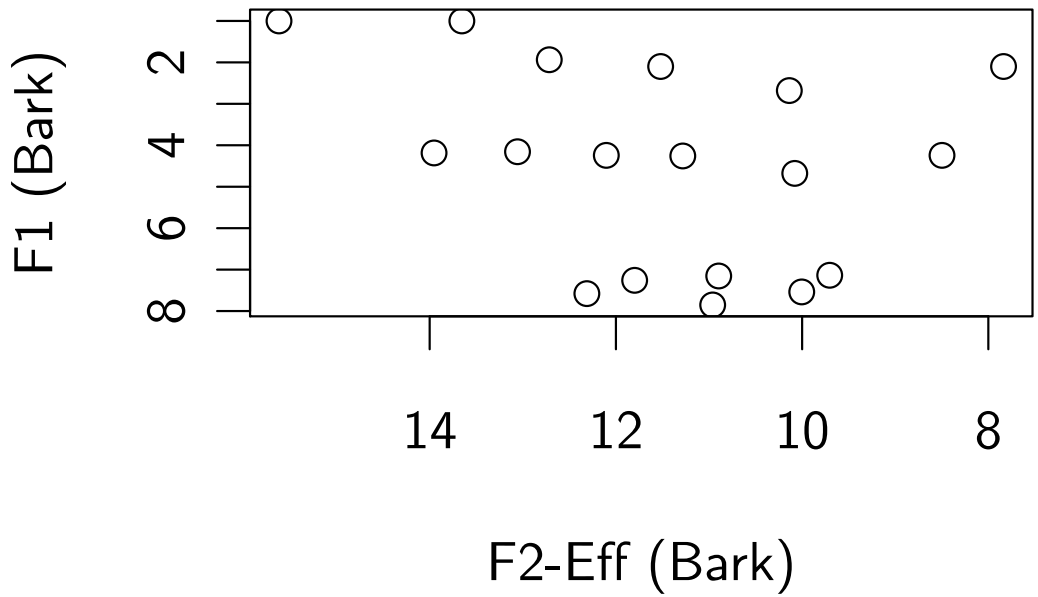
Articulatory synthesiser

- The synthesiser is based on interpolation (quadratic in the dimensions of height and position and linear in the dimension of lip rounding) between the formant frequencies of 18 artificially generated vowels.

Articulatory synthesiser

- $F1 = ((-392 + 392r)h^2 + (596 - 668r)h + (-146 + 166r))p^2 + ((348 - 348r)h^2 + (-494 + 606r)h + (141 - 175r))p + ((340 - 72r)h^2 + (-796 + 108r)h + (708 - 38r))$
- $F2 = ((-1200 + 1208r)h^2 + (1320 - 1328r)h + (118 - 158r))p^2 + ((1864 - 1488r)h^2 + (-2644 + 1510r)h + (-561 + 221r))p + ((-670 + 490r)h^2 + (1355 - 697r)h + (1517 - 117r))$
- $F3 = ((604 - 604r)h^2 + (1038 - 1178r)h + (246 + 566r))p^2 + ((-1150 + 1262r)h^2 + (-1443 + 1313r)h + (-317 - 483r))p + ((1130 - 836r)h^2 + (-315 + 44r)h + (2427 - 127r))$
- $F4 = ((-1120 + 16r)h^2 + (1696 - 180r)h + (500 + 522r))p^2 + ((-140 + 240r)h^2 + (-578 + 214r)h + (-692 - 419r))p + ((1480 - 602r)h^2 + (-1220 + 289r)h + (3678 - 178r))$

Articulatory synthesiser: Data Points[5]



data points of F_1 and F_{2EFF} in a Bark vowel space

[5] Adapted from Bart de Boer 2000

Articulatory synthesiser: Calculating F2'

- From *Hertz* to *Bark*
  - $Bark = \begin{cases} \frac{ln(Hertz/271.32)}{0.1719} + 2 & Hertz > 271.32 \\ \frac{Hertz-51}{110} & Hertz \leq 271.32 \end{cases}$

Articulatory synthesiser: Calculating F2'[6]

- Calculating $F_2'(== F_2eff)$
  - $$F2' = \begin{cases} F_2 & \text{if } F_3 - F_2 > c \\ \frac{(2-w_1)F_2+w_1F_3}{2} & \text{if } F_3 - F_2 \leq c \text{ and } F_4 - F_2 > c \\ \frac{w_2F_2+(2-w_2)F_3}{2} - 1 & \text{if } F_4 - F_2 \leq c \text{ and } F_3 - F_2 < F_4 - F_3 \\ \frac{(2+w_2)F_3-w_2F_4}{2} - 1 & \text{if } F_4 - F_2 \leq c \text{ and } F_3 - F_2 \geq F_4 - F_3 \end{cases}$$
- Where $c = 3.5$ Bark (the critical distance)
- And
  - $w_1 = \frac{c-(F_3-F_2)}{c}$
  - $w_2 = \frac{(F_4-F_3)-(F_3-F_2)}{F_4-F_2}$

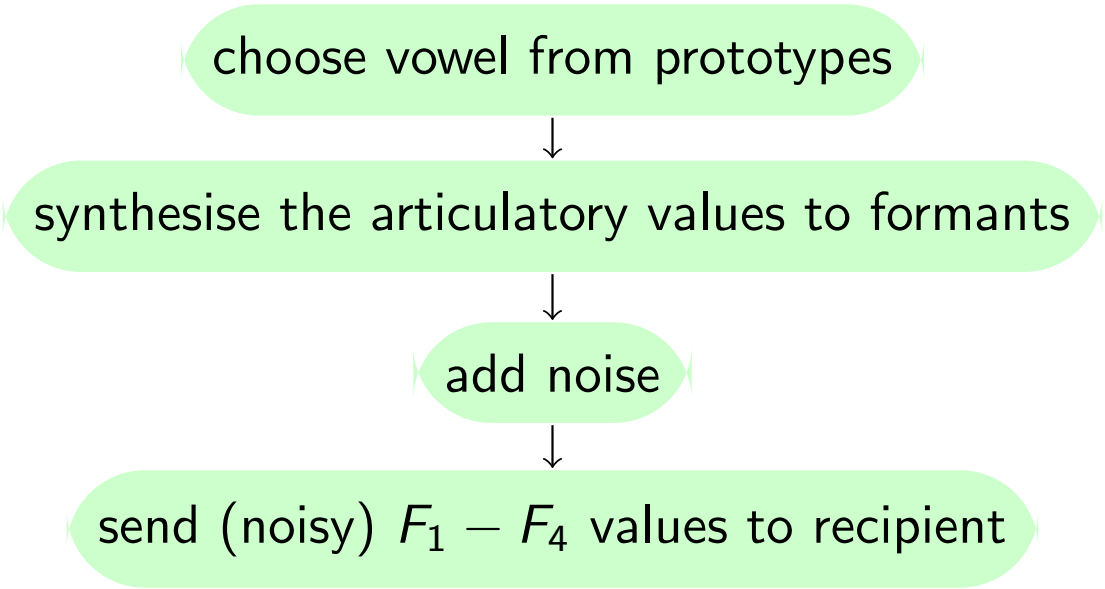
[6] Method by: Mantakas, Schartz & Escudier (1986)

Add noise to signal

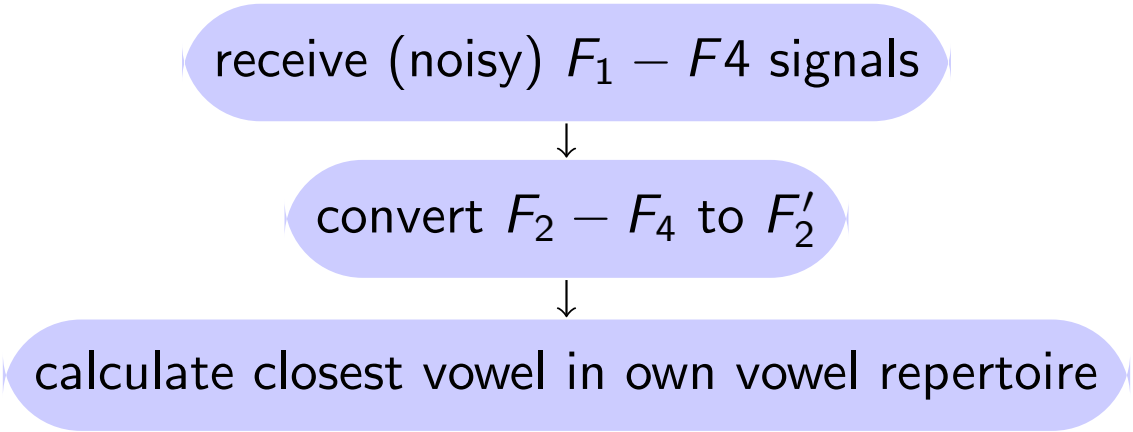
$$F_i' = F_i(1 + v_i)$$

- F_i' is the frequency of this formant after shifting and v_i is the shifting factor which is randomly chosen from the uniform distribution in the range $-\psi_i/2 \leq v_i < \psi_{ac}/2$, where ψ_{ac} is the maximal noise allowed, a very important parameter of the simulation. ψ_{ac} is a fixed parameter of the simulation.

The production chain



The perception chain



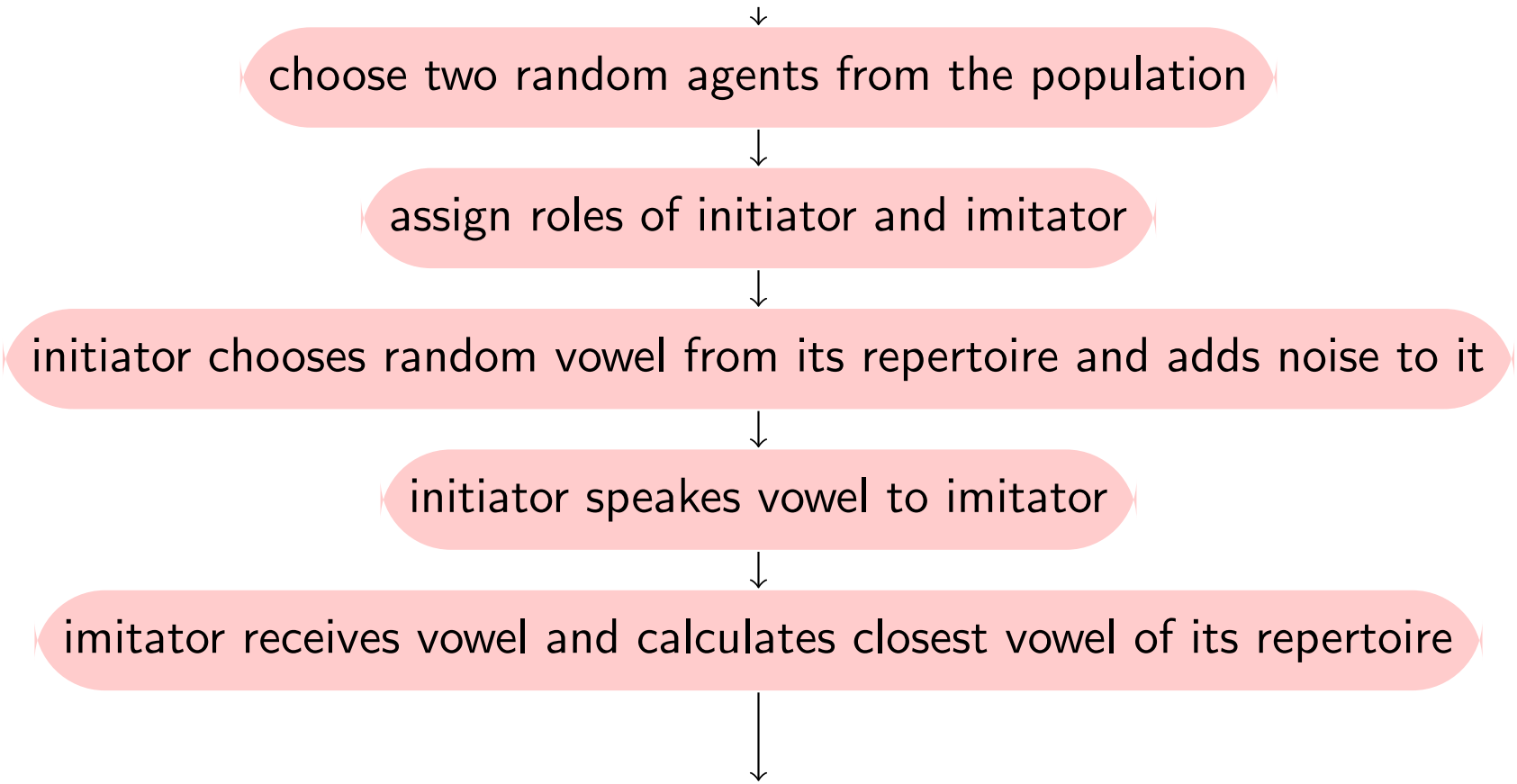
How are vowels perceived?

- The distance between two signals a and b is calculated by:
  - $D = \sqrt{(F_1^a - F_1^b)^2 + \lambda(F_2^{a'} - F_2^{a'})^2}$
- This is the weighted Euclidian distance between two vowels in the F_1 and F_2' space

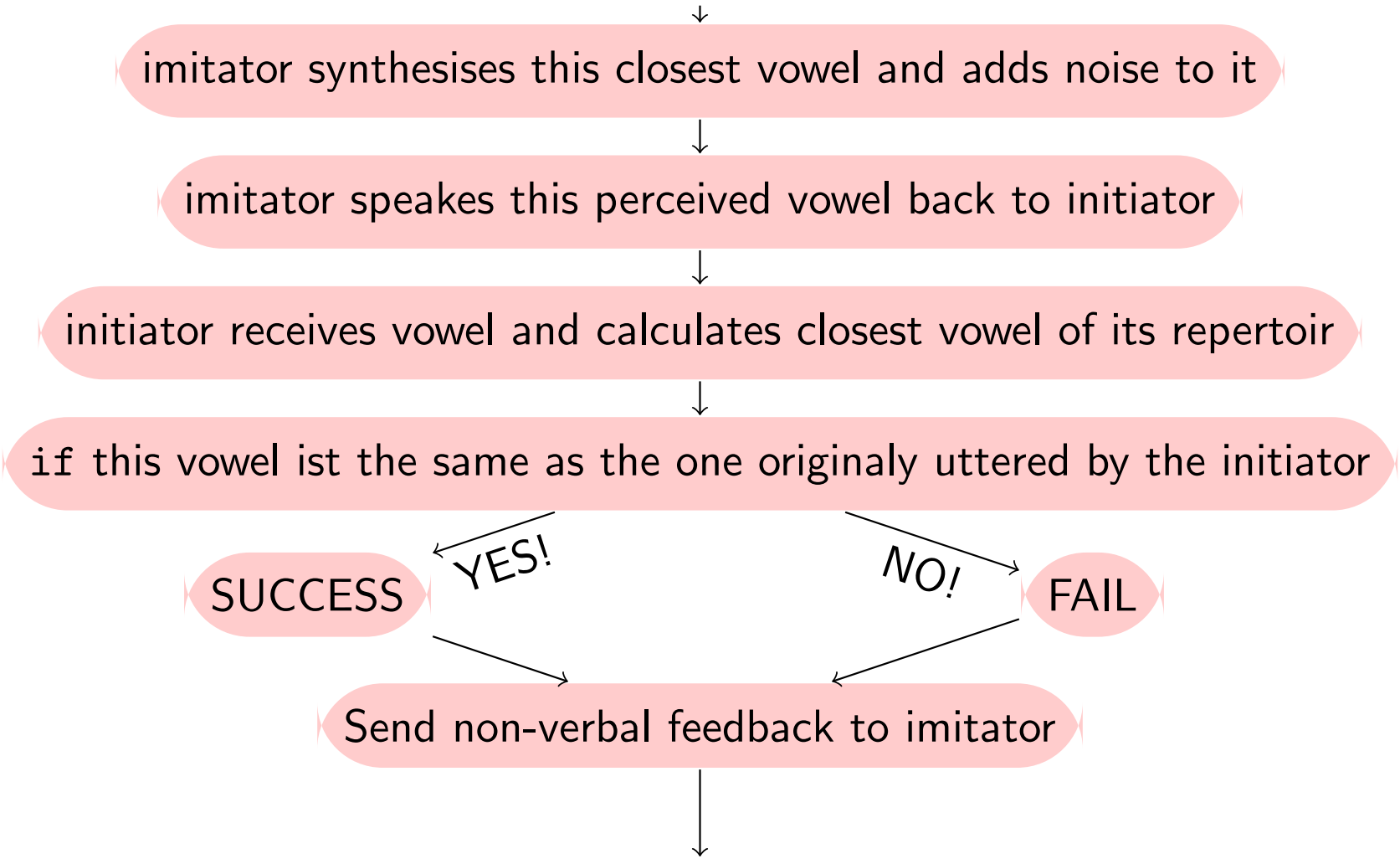
The imitation game

- Definition:
  - Imitation games are played between two agents whose goal is to imitate the other agent as well as possible.

The imitation game



The imitation game



The imitation game

↓

update repertoires according to the success of the game

↓

- Both the imitator and the initiator keep track of the number of times each of their vowels has been used and the number of times it has been used in successful imitation games.
- In case of successful imitation game:
  - Imitator moves vowel closer to the one received
- In case of unsuccessful imitation game:
  - If used vowel has a high success/use ratio → add vowel to middle of the articulatory space (all parameters set to 0.5) then iteratively shift it closer to the perceived vowel.
  - If used vowel was unsuccessful in other imitation games → shift towards observed acoustic signal

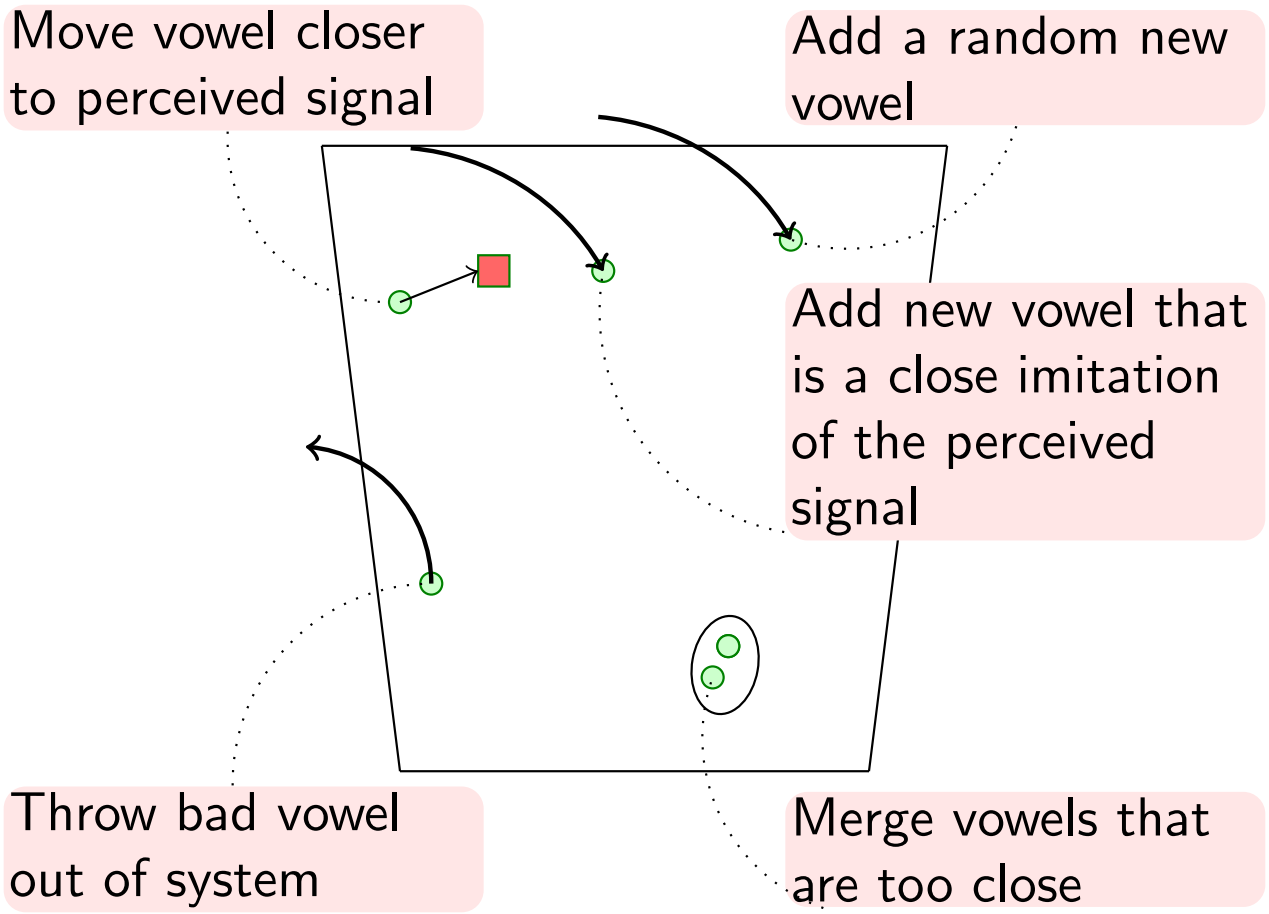
The imitation game

↓

perform repertoire updates

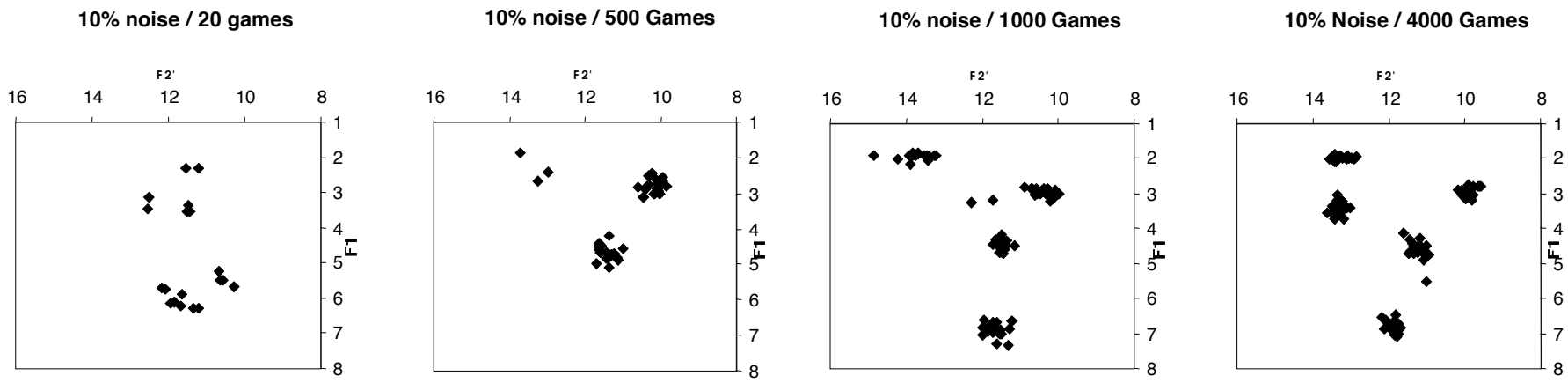
- Clean up vowel systems (remove ones that where not successful enough)(with probability p_c after every game)
- Clusters that could be noise are merged into a single vowel (best agent is chosen)
- Add new random vowel (with a low probability of p_i after every game)

The imitation game[7]



[7] adapted from Bart de Boer 2000

Results[8]



Development of a vowel system

[8] from Bart de Boer 2000

Possible expansion of the system

- **Problem/Limitation:**
  - Non-dynamic modelling of vowels!
- A possible expansion to start modelling formant movements:
  - Synthesise start and end points of vowel movements
  - Interpolate them
  - Send line/curve information to imitator
  - Calculate correlation to existing movements in repertoire
  - Adjust line/curve (slope/y-offset) due to imitation game outcome

Possible expansion of the system

Thank you for your attention!
Questions???