The dynamical approach to speech perception: from fine phonetic detail to abstract phonological categories

Noël Nguyen

Laboratoire Parole et Langage
CNRS & Université de Provence
Aix-en-Provence, France

Much attention has been devoted recently to the potential role of phonetic detail in the perception and understanding of speech

This role is minimized by abstractionist models in which speech is mapped onto context-independent abstract features

However, recent research suggests that listeners are sensitive to phonetic detail and that lexical representations contain fine phonetic information

In this talk, the opposition between abstractionist and exemplar models of speech perception will be discussed

I will also offer new empirical evidence for a non-linear dynamical model of speech perception (Tuller and coll., 1994) in which perceptual categories are associated with attractors of a potential function

The abstractionist approach

- Concept of perceptual normalization
- Underlying phonological representations are abstract, discrete and context-free
- Partially arbitrary semantic relationship between phonological representations and phonetic forms (eg concept of phonetic interpretation in Firthian Prosodic Analysis, Local, 1992)
- Emphasis on the mental lexicon as a set of permanently stored, context-independent word units

The exemplar-based approach

- Representation-based approach to the speech variability problem (Pitt & Johnson, 2003)

- Words and high-frequency grammatical constructions stored in memory as lists of exemplars

- Exemplars are highly context-dependent; they contain fine-grained phonetic detail that conveys both indexical and linguistic information

(Bybee, 2001; Coleman, 2002; Elman, 1995; Docherty, 2003; Goldinger, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert, 2002, inter alia)

- Alternative to combinatorial paradigm (Bybee & McClelland, 2005)
- In some models at least, exemplars have no internal structure and are unanalyzed auditory representations (eg Hawkins, 2003; Johnson, 1997)
- However, phonological units such as segments and syllables may be brought to the listener's consciousness as the speech signal is mapped onto the lexicon
- These units are a temporary by-product of lexical activation, and they emerge as connections between time-aligned, phonetically-similar portions of exemplars are established
- There is no basic unit of speech perception: units of different sizes may be simultaneously activated, with a natural bias for larger units to prevail upon smaller ones (Goldinger & Azuma, 2003; Grossberg & Myers, 2000)

The exemplar-based approach (cont.)

- Phonetic/phonological knowledge includes *both* abstract patterns (eg CVC schemas) and token-specific detail (Langacker, 2000)

- Concept of *phonetic similarity* is central

  In speech understanding, phonetic similarity determines the pattern of activation in the lexical space as well as the emergence of sublexical units

- Frequency of use also has a major role in perception (eg more frequent phonetic features resonate with the input before less frequent ones, McLennan & Luce, 2005), and has an effect on how words and constructions are represented in memory

Exemplar models: three assumptions

- Fine phonetic detail has a direct influence on patterns of lexical activation
- Emergent segmental units are based on auditory similarity between overlapping portions of exemplars and are therefore context-dependent
- These units, however general they may be, arise from the sounds listeners are overtly exposed to; no role assigned to abstract phonological entities such as *empty onsets* or *floating segments* for example

Role of fine phonetic detail in speech understanding

- Listeners are sensitive to fine-grained phonetic cues in speech perception and word recognition: subphonemic variations in VOT in syllable-initial stops, V-to-V coarticulatory patterns, long-domain resonance effects associated with liquids, long-domain acoustic cues to coda voicing, etc.
  (see Hawkins, 2003, for a review)

- Prior exposure to an utterance facilitates later recognition (Goldinger, 1996)

Exemplar models and allophonic variation

- Emergent segmental units are based on auditory similarity between overlapping portions of exemplars and are therefore context-dependent

- Bybee, 2001: Determination of how to categorize a phonetic segment is based on its substantive properties and not on its distribution

  Phonetic tokens are classified as members of the same category if they are highly similar in their acoustic/articulatory properties

  Example: the durational difference between American English stop [d] and flap [ɾ] is large enough to require a separate category for the flap

- McLennan, Luce & Luce (2003) used long-term repetition priming to determine whether flaps are represented veridically as opposed to being mapped onto underlying phonemic units

  Their results are not entirely consistent with either abstractionist or exemplar models

- Pegg & Werker (1997) and Whalen et al. (1997) found that allophonic variants are more difficult to discriminate than phonemic contrasts

- Peperkamp et al. (2003) provide evidence suggesting that the difference between syllable-final uvular voiced fricative [ʁ] vs voiceless [χ] (both allophones of /r/ in French) is more difficult to perceive than that between phonemes /m/ and /n/, in the context of a following CV sequence

Nguyen, Dufour, Frauenfelder & Meunier (2005): Perception of allophonic variations in mid vowels of Southern French

- Northern French: contrastive distinction between /e/-/ɛ/, /ø/-/œ/ and /o/-/ɔ/, e.g.

été	[ete]	*saute*	[sot]
étais	[etɛ]	*sotte*	[sɔt]

- Southern French: no contrastive distinction between /e/-/ɛ/, /ø/-/œ/ and /o/-/ɔ/; the distribution of the mid-high and mid-low variants is said to be entirely governed by a variant of the *loi de position* (the mid-high variant occurs in open syllables and the mid-low variant in closed syllables and whenever the next syllable contains a schwa, Durand, 1990), e.g.

été	[ete]	*saute*	[sɔtə]
étais	[ete]	*sotte*	[sɔtə]

How do speakers of Northern and Southern French perceive the e/ɛ and o/ɔ word-final contrasts in word recognition?

In a standard abstractionist model of speech perception, one may assume that minimal pairs ending in a mid-high vs mid-low vowel will be mapped onto a single underlying abstract phonological representation, and that both forms will be processed as being identical by Southern French listeners

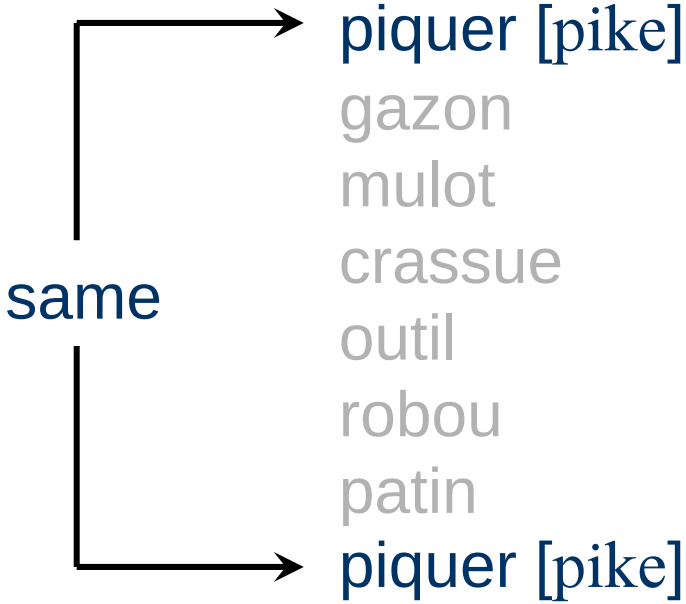
The repetition priming paradigm

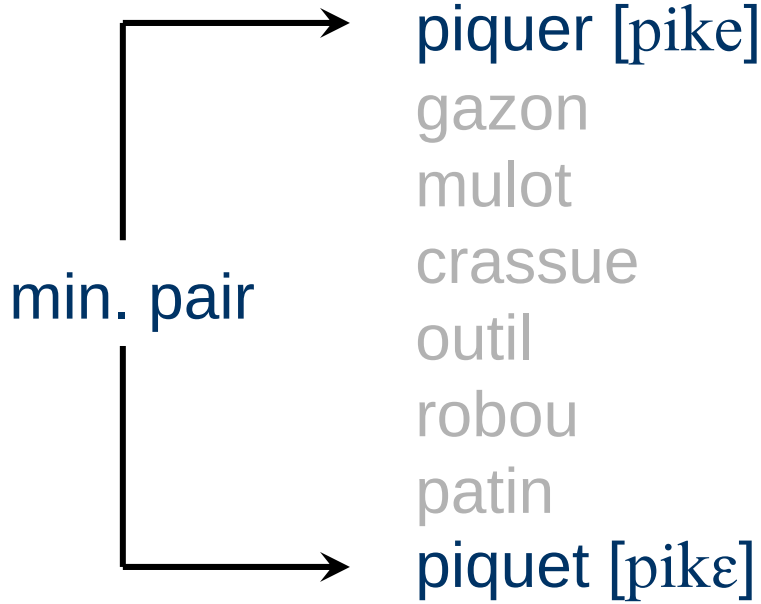
piquer → RT1
gazon
mulot
crassue
outil
robou
patin
piquer → RT2

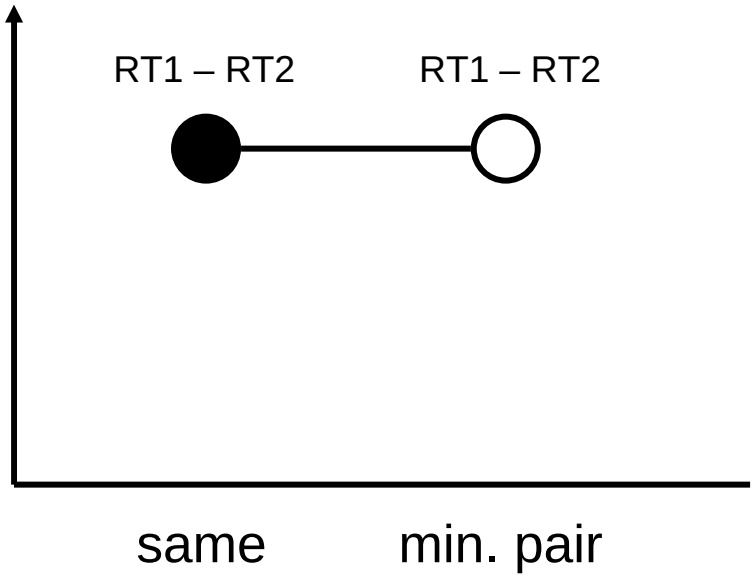
RT2 < RT1

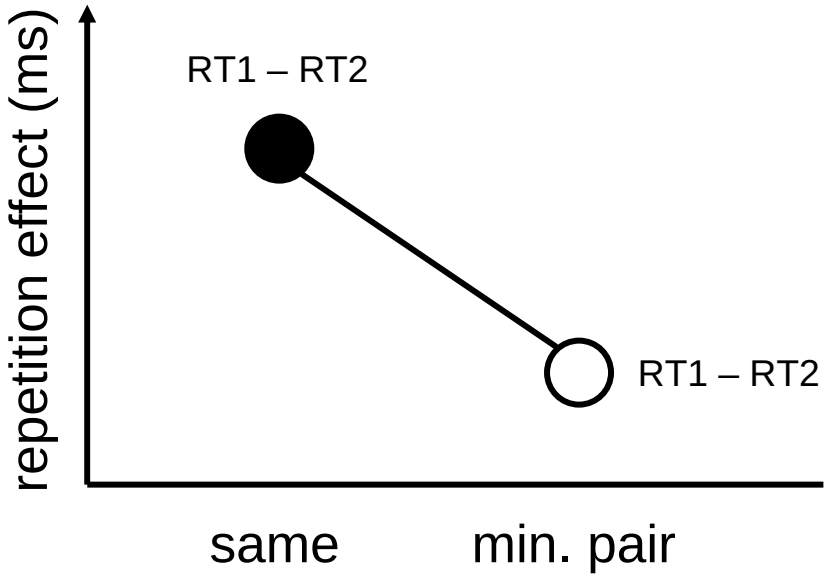
feuquer → RT1
gazon
mulot
crassue
outil
robou
patin
feuquer → RT2

RT2 ≈ RT1


piquer [pike]
gazon
mulot
crassue
same
outil
robou
patin
piquer [pike]


piquer [pike]
gazon
mulot
crassue
min. pair
outil
robou
patin
piquet [pikɛ]





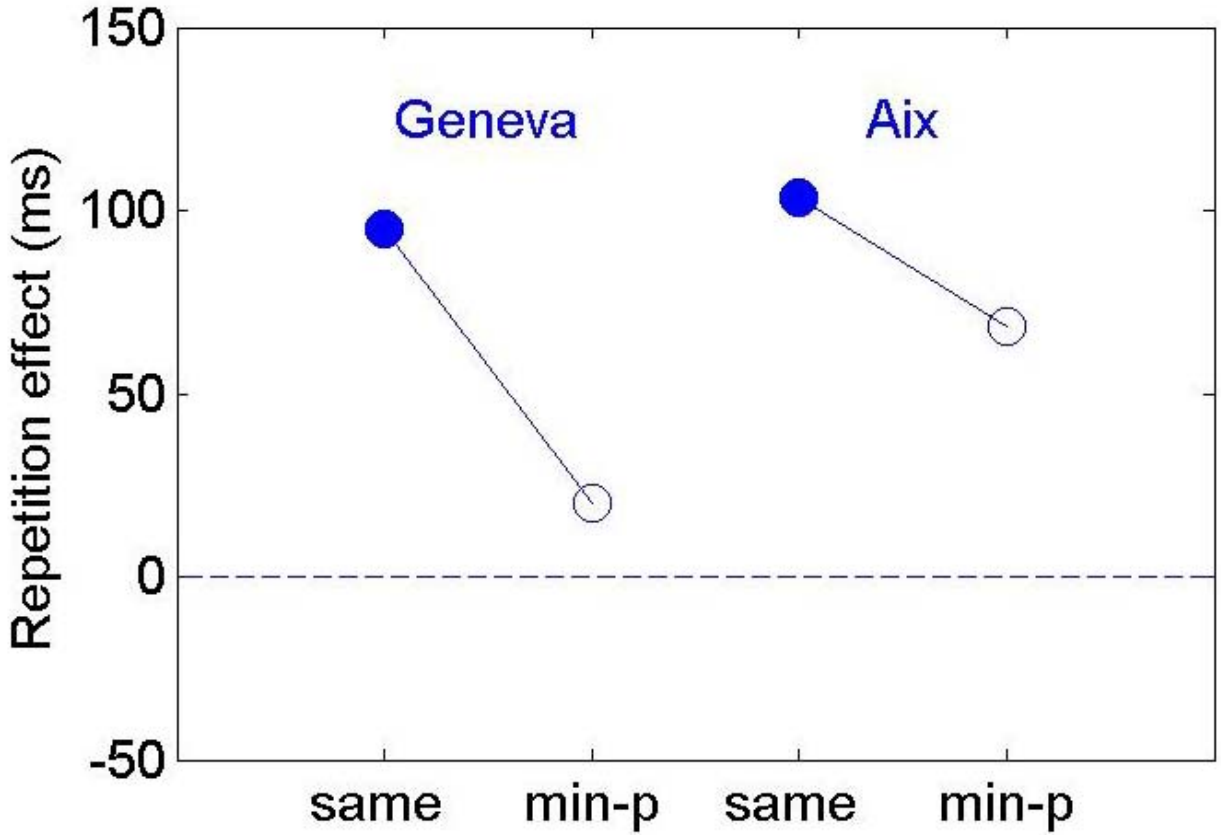
No decrease of priming effect for minimal pairs relative to identical pairs:

⇒ prime and target are mapped onto same underlying representation

Decrease of priming effect for minimal pairs relative to identical pairs:

⇒ prime and target are not associated with same underlying representation

words ending in /e/ vs /ɛ/



First results suggest that Southern French listeners are sensitive to word-final e/ɛ and o/ɔ contrasts in word recognition, albeit to a lesser degree than Northern French listeners

Exemplar models and abstract phonological entities

In exemplar models, emergent segmental units, however general they may be, arise from the sounds listeners are overtly exposed to; no role assigned to abstract phonological entities such as *empty onsets* or *floating segments* for example

Nguyen, Wauquier-Gravelines, Lancia & Tuller (2005) have examined this assumption in an investigation on the perception of liaison in French

Liaison in French

Liaison: appearance of a consonant (liaison consonant, LC) at the juncture of two words, which otherwise are not pronounced with that consonant

→ Word1 – LC – Word 2

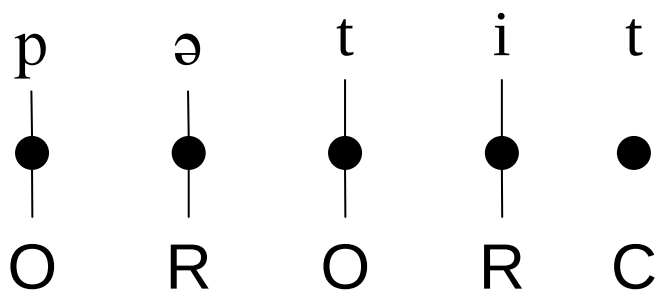
examples:

- petit ours [pətituʁs] « small bear »
- les amis [lezami] « the friends »
- vous allez [vuzale] « you go »
- en avant [ɑ̃navɑ̃] « in front »

Two phonological accounts of liaison

- Autosegmental approach (Encrevé, 1988)
- Exemplar-based approach (Bybee, 2001)

The autosegmental account (Encrevé, 1988)

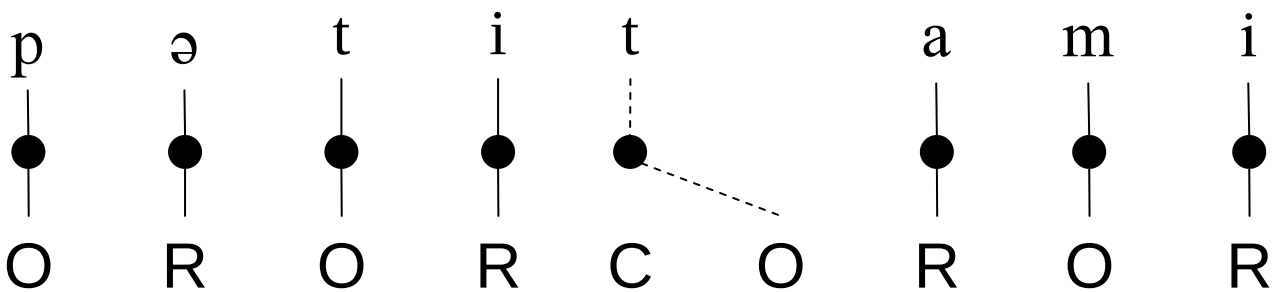


segmental tier

skeletal tier

syllabic tier

The liaison consonant is floating with respect to both the skeletal and syllabic tiers



A skeletal slot is available that allows the anchoring of LC both to the skeleton and to the syllable tier

The exemplar-based account (Bybee, 2001)

- Liaison occurs within grammatical constructions

  ex.: NOUN – z – [vowel]-ADJ_{Plural}

- Grammatical constructions range on a continuum from the very general (see above) to the very specific (e.g. *c'est-à-dire*); this accounts both for false liaisons (overgeneralization of a construction, ex.: *chemins de fer* [z] *anglais*) and word-specific differences in the realization of liaison

- Grammatical constructions are both storage and processing units

Importantly, liaison consonants do not have a specific status relative to that of the other segments of the construction, in the exemplar-based approach. They are entrenched in the construction and belong to the same plane as the segmental units in the preceding and following words.

In the autosegmental approach, by contrast, the characterization of liaison consonants as floating segments provides them with a highly specific status.

Consequently, one issue addressed in the present work is whether liaison consonants are processed in the same way as non-liaison consonants.

Experimental design (based on Wauquier-Gravelines, 1996):

Phoneme detection task (/n/ or /z/)

The target consonant can appear:

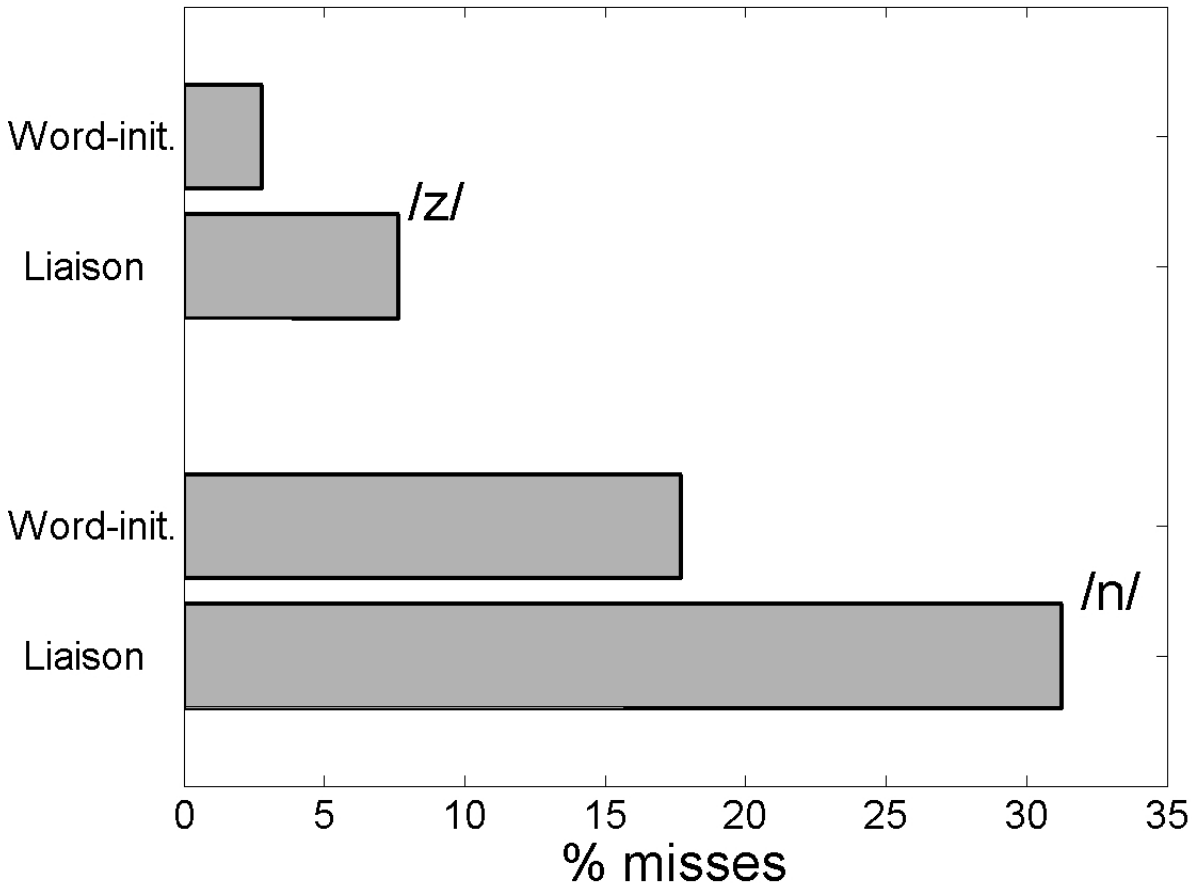
- in word-initial position, e.g.:

  Il dépasse un nageur [œ̃naʒœʁ]…

- in liaison position, e.g. :

  Elle repasse un habit [œ̃nabi]…

Results: proportion of misses



The target consonant is more difficult to detect in liaison position than in word-initial position

The response patterns do not provide support for the exemplar-based approach. In this approach, the liaison consonant does not have a specific status compared to the other segments in constructions. Under the exemplar-based account, liaison /n/ should have been in fact *easier* to detect than word-initial /n/, since the former is more frequent than the latter in the contexts we used.

The results are in better agreement with the autosegmental model. In this model, liaison consonants are *structurally unstable*. The data suggest that indeed liaison consonants may not have the same phonological status as fixed consonants for the listener.

Beyond the abstract representations vs exemplars dichotomy?

Exemplar models:

- Account for listeners' sensitivity to fine phonetic detail, indexical variation, frequency of occurrence
- Provide an alternative to the normalization hypothesis
- Emphasize the links between speech perception and other forms of perceptual categorization
- Provide an explanation for how phonological categories may emerge and show that the « basic unit of speech perception » may be an ill-posed problem

Exemplar models, however…

- Do not seem to be able to fully account for how allophonic variation is dealt with by listeners (e.g. Peperkamp et al., 2003)

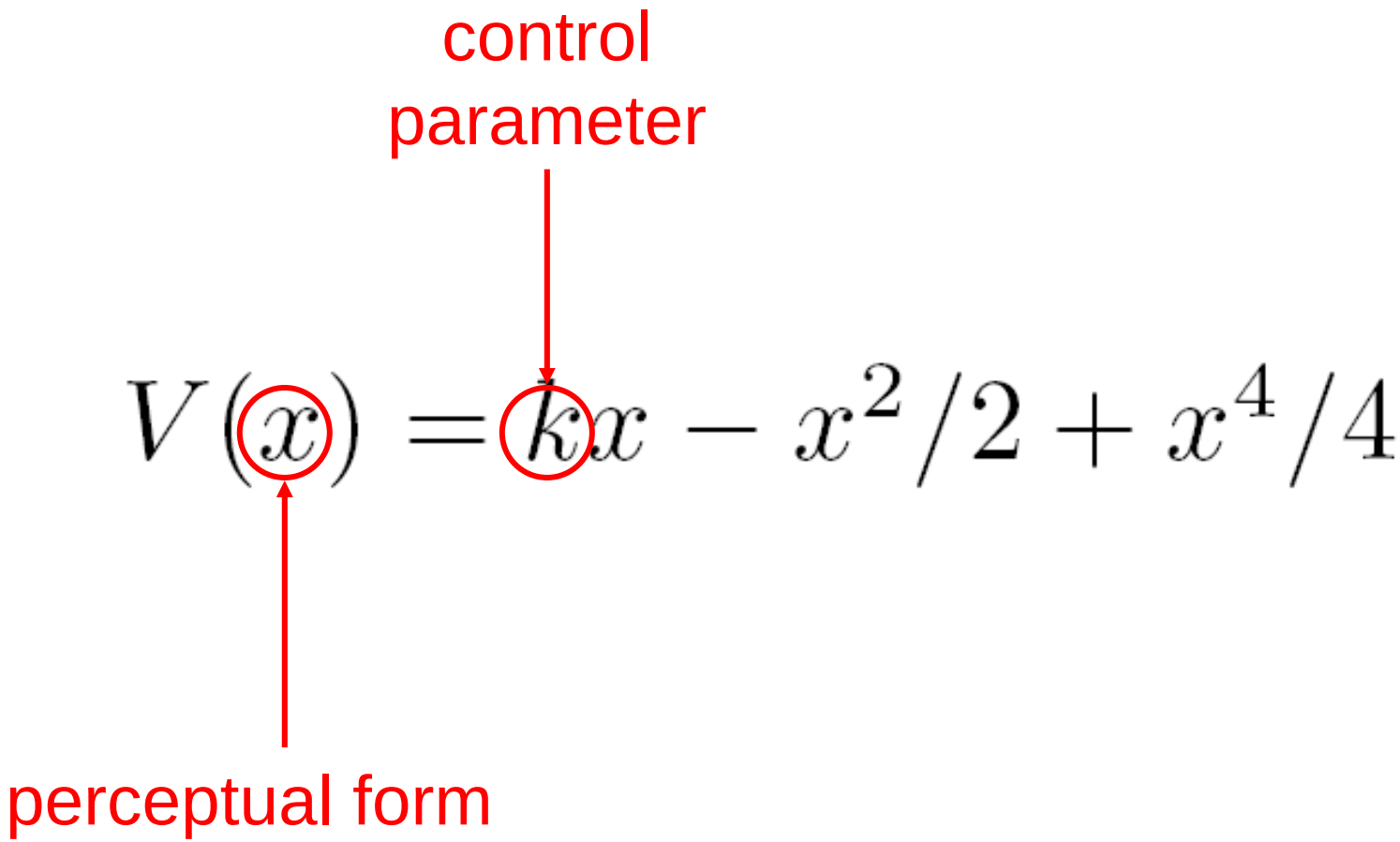
- More generally, may put too strong an emphasis on the role of auditory similarity and inductive generalization in the emergence of phonological categories

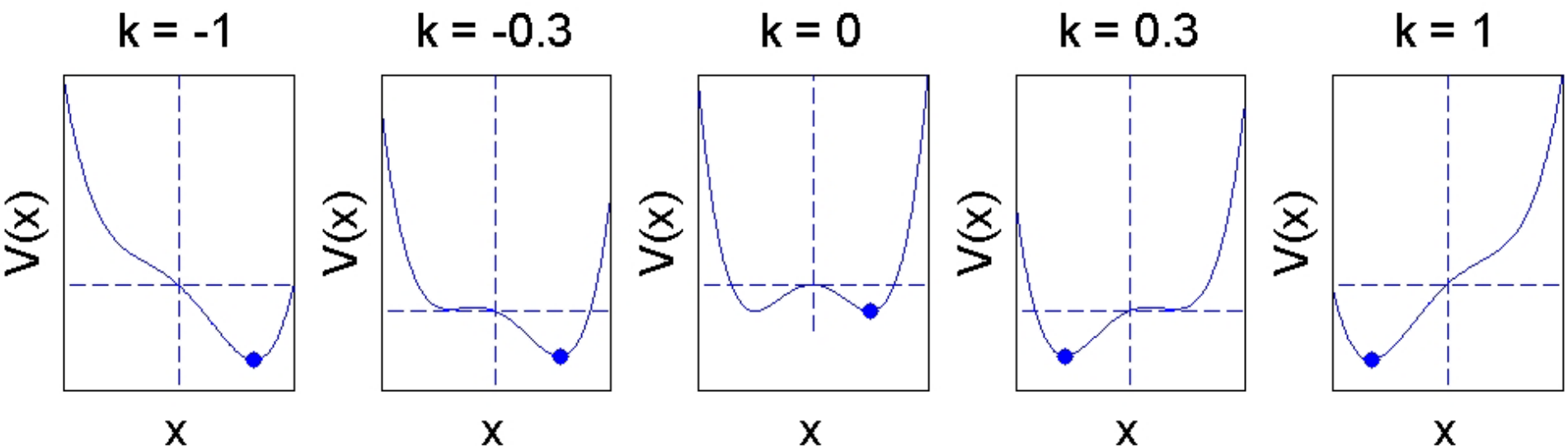
- Fail to explain why, in certain circumstances, listeners seem insensitive to variations in the surface forms of words (eg Lahiri, 2005; Pallier et al., 2001) and why listeners find it difficult to detect high-frequency liaison consonants in the speech chain (Wauquier-Gravelines, 1996; Nguyen, Wauquier-Gravelines, Lancia & Tuller, 2005)

Towards a dynamical approach

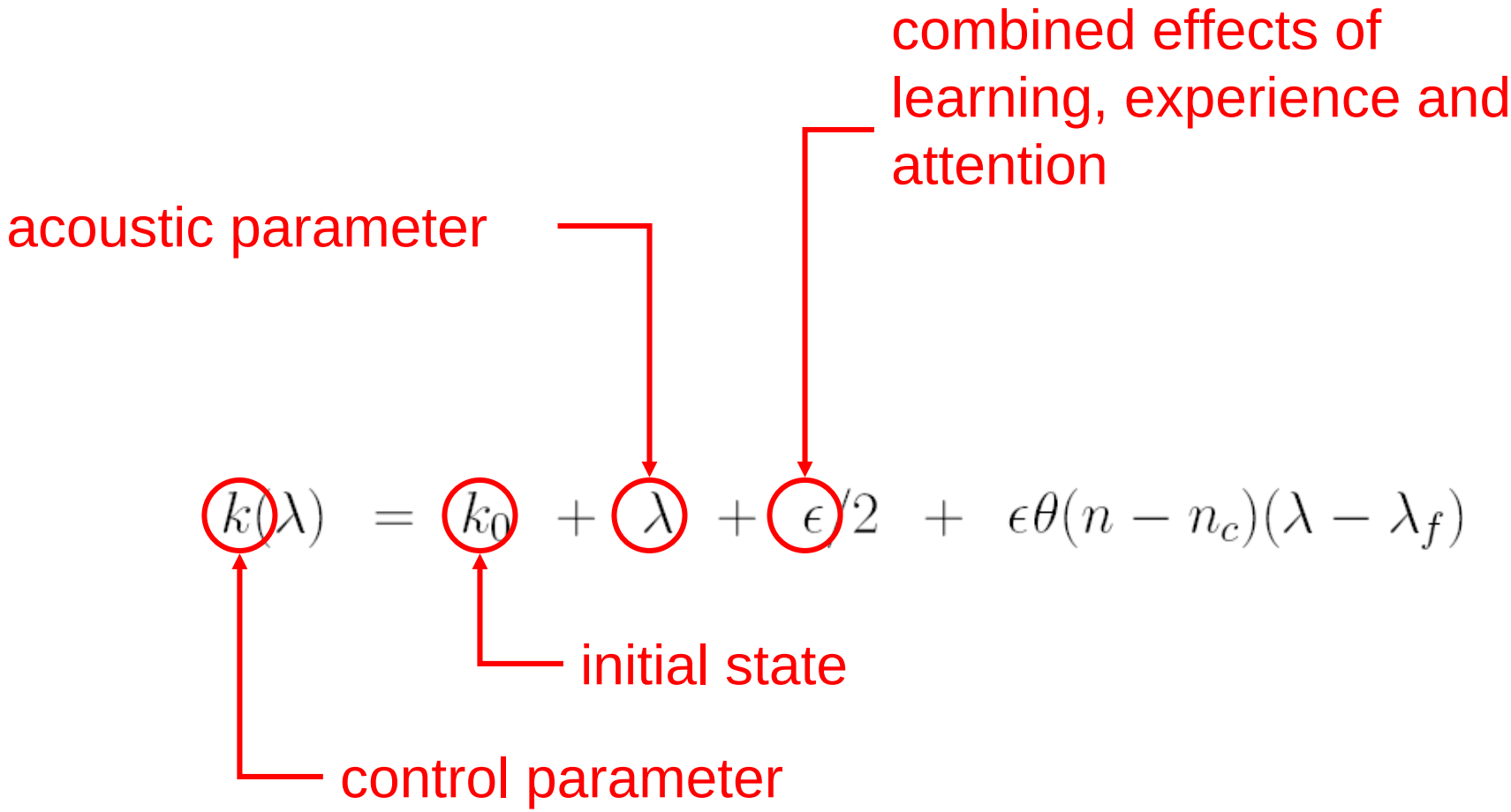
Tuller, Case, Ding & Kelso (1994):

- The perceptual system is a non-linear dynamical system, whose behavior depends on its previous state as well as a number of control parameters
- Perceptual categories are associated with attractors of a potential function
- The system's behavior can show qualitative changes over time under the influence of the control parameters (eg abrupt shift toward another attractor)
- The availability of a percept, its stability and strength, are functions of the acoustic properties of the stimulus, the previous percept, and the combined effects of learning, linguistic experience and attentional factors


control
parameter
$V(x) = kx - x^2/2 + x^4/4$
perceptual form



Potential landscape for five values of control parameter k (after Tuller et al., 1994)


combined effects of learning, experience and attention
acoustic parameter
$k(\lambda) = k_0 + \lambda + \epsilon/2 + \epsilon\theta(n - n_c)(\lambda - \lambda_f)$
initial state
control parameter

Tuller and colleagues investigated the perceptual dynamics of speech categorization when the stimuli are presented *sequentially* along a relevant acoustic dimension

Stimuli ranging on a *say-stay* continuum were presented to listeners in a sequential order (e.g. from *say* to *stay* and back to *say*, by incrementally increasing then decreasing the duration of the silent interval between /s/ and /eɪ/)

The response patterns showed a number of dynamical characteristics which included:

- *Hysteresis* (listener's initial response tends to persist across the continuum)
- *Enhanced contrast* (listener quickly switches to alternate percept and does not hold on to initial categorization)
- *Critical boundary* (switch between percepts remains associated with the same stimulus regardless of presentation order)

Nguyen, Bergounioux, Lancia, Wauquier-Gravelines and Tuller (2005), extended this experimental paradigm to French and explored the role of long-term training on categorization.

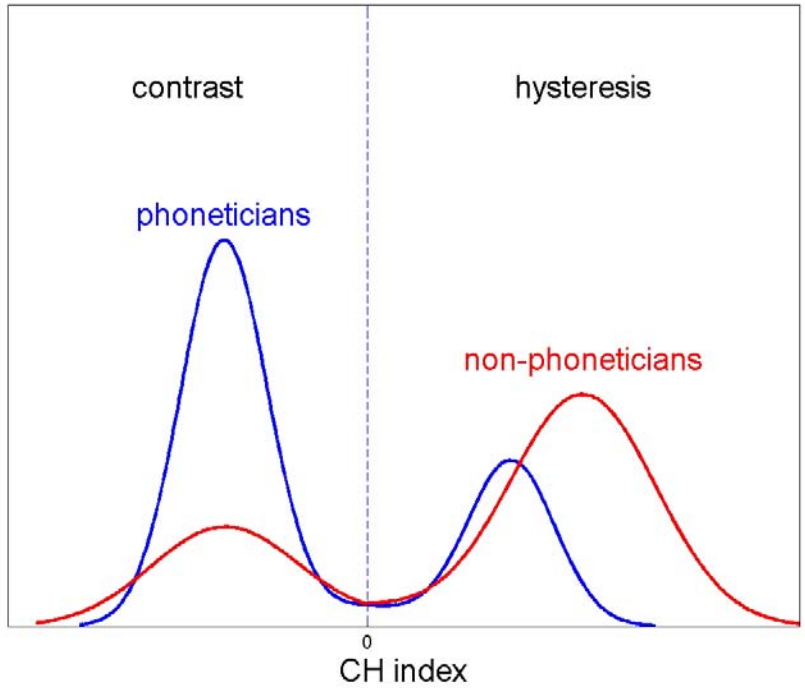
Their goal was to determine to what extent training has an influence on the stability of percepts, and on the dynamical characteristics of categorization.

The experiments revealed that:

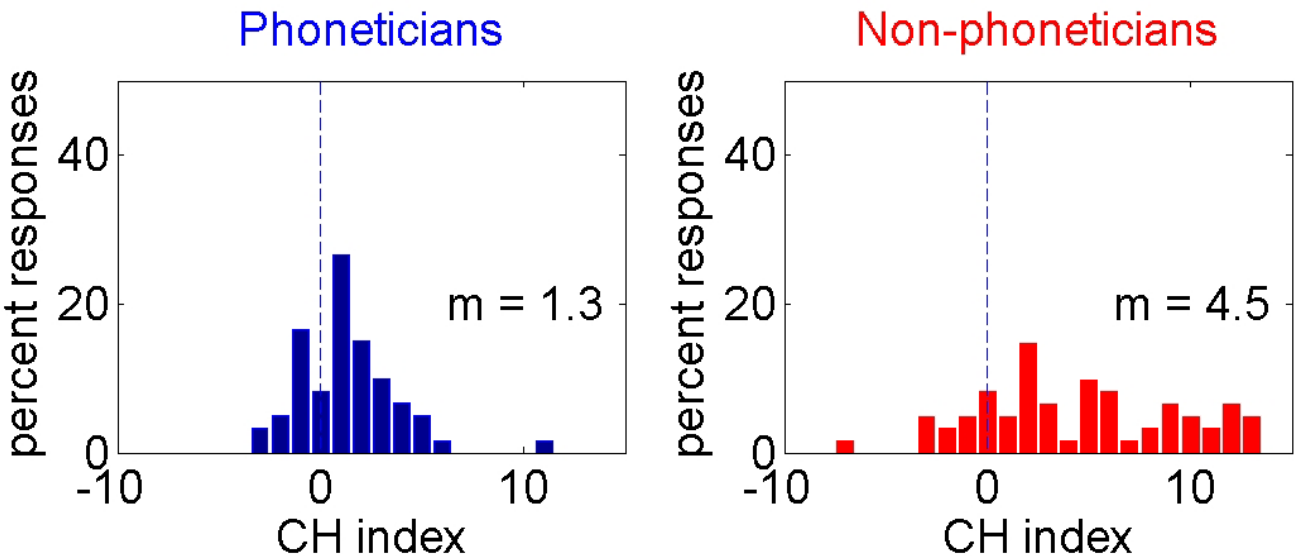
- Hysteresis was the most frequent pattern, followed by contrast, with critical boundary occurring least often
- Untrained listeners showed more hysteresis and less contrast than trained listeners
- Untrained listeners showed more perceptual instability than trained listeners

These findings were consistent with the model's predictions

Predicted relationship between contrast and hysteresis depending on phonetic training



Observed response patterns



Discussion

- The Tuller et al. model displays a number of desirable properties that are also shared by exemplar models, such as sensitivity to fine-grained phonetic detail and to frequency of occurrence, and attunement to the speaker's individual characteristics

- Unlike exemplar models, however, the Tuller et al. model does not posit that perceptual categories are isomorphic to auditory speech patterns. Whereas the acoustic characteristics of the stimulus have an influence on the shape of the potential function, this influence is conveyed through a non-linear function, and combined with high-level cognitive factors such as attention, experience and training

- Attractors associated with the potential function can be viewed as a discretization of the perceptual space

- However, the potential function itself is continuous, and so is the the sound-to-percept mapping (see Gafos, 2004)

Next step: modelling the perception of liaison consonants in a dynamical framework

Merci