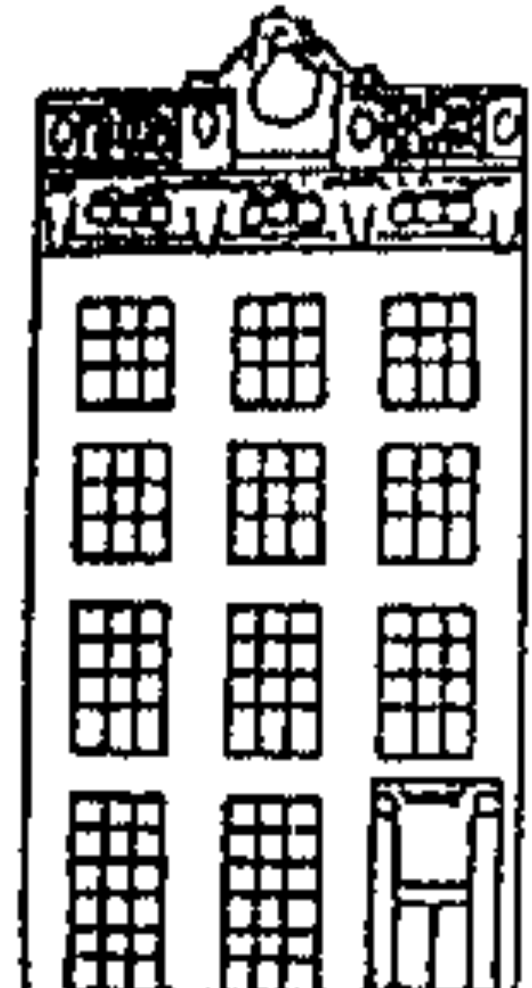
PHONEME RECOGNITION AS A FUNCTION OF TASK AND CONTEXT

R.J.J.H. van Son and Louis C.W. Pols
Institute of Phonetic Sciences IFA/ACLC
University of Amsterdam, The Netherlands
Rob.van.Son@hum.uva.nl



Introduction

Phoneme recognition has 2 meanings**:**

1 Phoneme naming

2 Phone categorization

Where phone categorization precedes phoneme naming

Ad 1: **Phoneme naming**

- Consious	(identification)
- Lexical	(results in a label)
- Competitive	(winner takes all)
- Prime-able	
- Frequency sensitive	

Ad 2: **Phone categorization** (hypothetical)

- Pre-consious / 'On-line'
- Pre-lexical
- Many categories can be activated
- Unprime-able?
- Frequency effects are 'intricate'

Units of speech

Phones or Phonemes are considered here to be the units of speech (which is an over-simplification)

Definition of Phonemes:
Smallest "unit of difference" between words.
Phonemes are described as **Feature bundles**

Examples:			(feature difference)
[tEnt]	<-->	[dEnt]	(voicing)
[tEnt]	<-->	[kEnt]	(place of articulation)
[dEnt]	<-->	[kEnt]	(both)

Not all possible feature bundles are legal phonemes:
> 600 phonemes known worldwide
English uses **< 50**

Every language differs in the way it defines features

Example: Voicing

English	/tEnt/ & /dEnt/	--> Dutch	[tEnt]
Dutch	/tEnt/ & /dEnt/	--> English	[dEnt]

Phonotactics

Not all phoneme combinations are legal

Phonotactic & phonological rules define
legal phoneme and feature combinations
These rules define the smallest possible
differences between words

Examples:
[tEnt] <--> [tEnd] (English, *Dutch)
[tEnt] <--> [tEnk] (*English, *Dutch)
[tEnt] <--> [tENK] (English, Dutch)

Phonotactic & phonological rules are a syntactic layer over the phoneme sets. Phoneme inventory and phonology are optimized with respect to each other.

Phonemes define legal feature combinations. Phonological (phonotactical) rules define legal feature sequences.

Both "illegal" phonemes and "illegal" phoneme sequences hamper production and perception

The role of Phonemes in speech recognition

Two opposite (extreme) hypotheses:

A) Obligatory phoneme hypothesis
All speech is converted to a string of phoneme symbols before lexical access.
Phoneme categorization is absolute and obligatory.

B) Lax phoneme hypothesis
Phonemes (or phones) are the result of prelexical regularization and data reduction processes that extract the relevant acoustic information.
The phone(me)s are clustering artefacts of the extracted information, i.e., they represent the prefered "acoustic events".
In this hypothesis, categorization is partial and can be defered.

What makes a phoneme?

Does every phoneme have a unified and unique canonical target?
(both in production and perception)

Unlikely cf.:
different phones/same phoneme
/l/ in [hOl] and [lOw] (dark vs light)
/t/ in [dEnt] and [tEnd] (unreleased vs aspirated)

same phones/different phonemes ("bad bet" exp)
/E/ vs/ae/ in [bEd] and [baet] (short vs long)
/d/ vs /t/ in [baed] and [bEt] (-/+ voiced)
/I/ vs /O/ in [mIljun] and [bijOsko:p] (Dutch)

A phone is a realization of a phoneme only in a certain context.
Allophones of a phoneme do not have to have anything in common at all.

Proposition:
The identity of a phone in context is completely at the discretion of the language and how it optimizes the trade-off between ease of production and perception.

The acoustics of phonemes

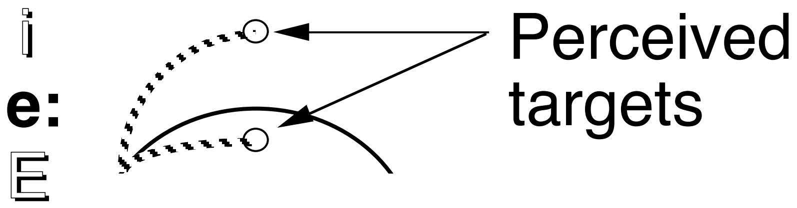
Classical approaches:

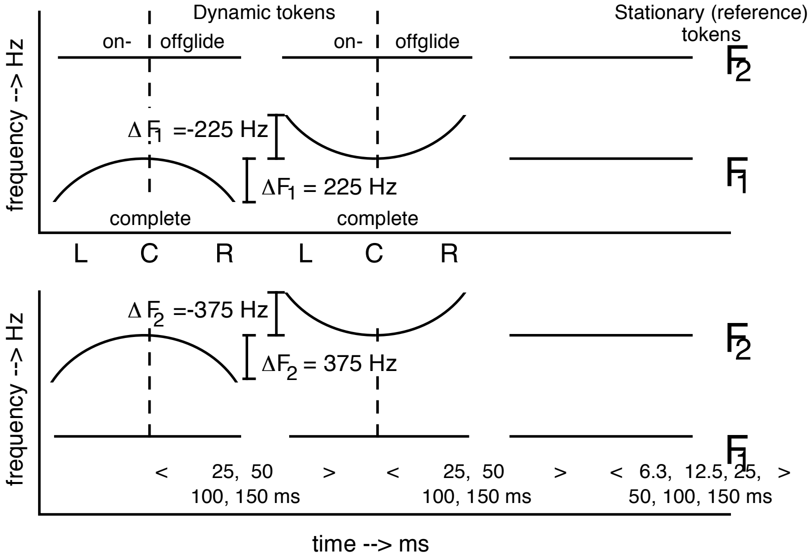
A) Static clustering theories
Each phoneme is a simple, continuous category in some perceptual space. Requires rather complex acoustic transformations (normalizations).

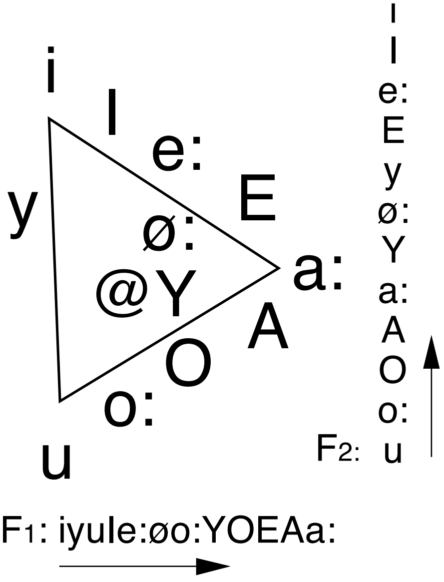
B) Dynamical specification
The dynamics of speech generates predictable deviations from the canonical target that can be undone by extrapolation of suitable parameter tracks or inverse modelling (motor theory).

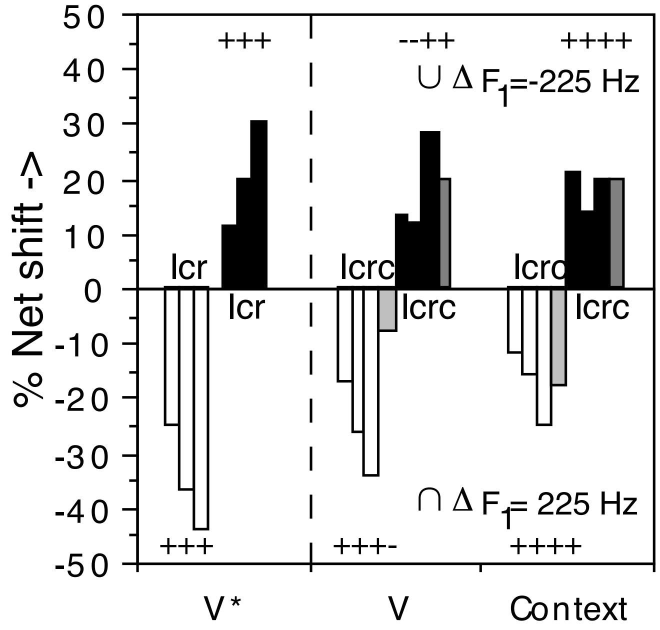
Both theories have problems with some data (proponents of both theories have thoroughly disproved each other's point).

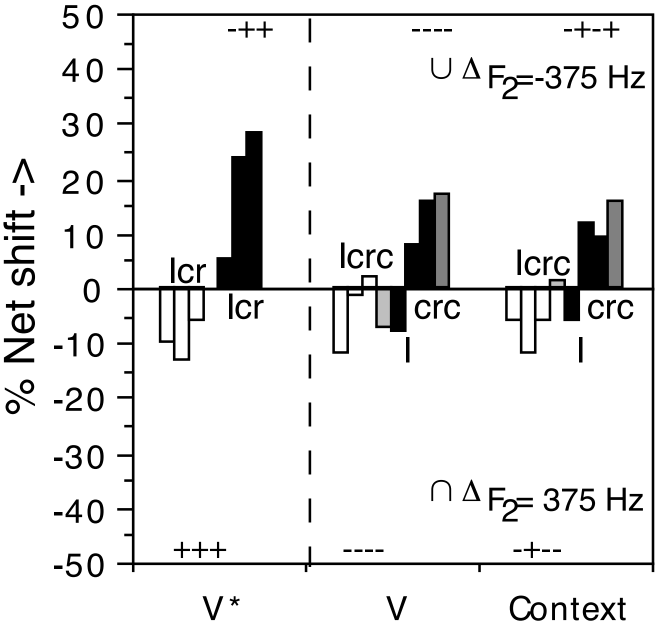
Example: Target overshoot ?











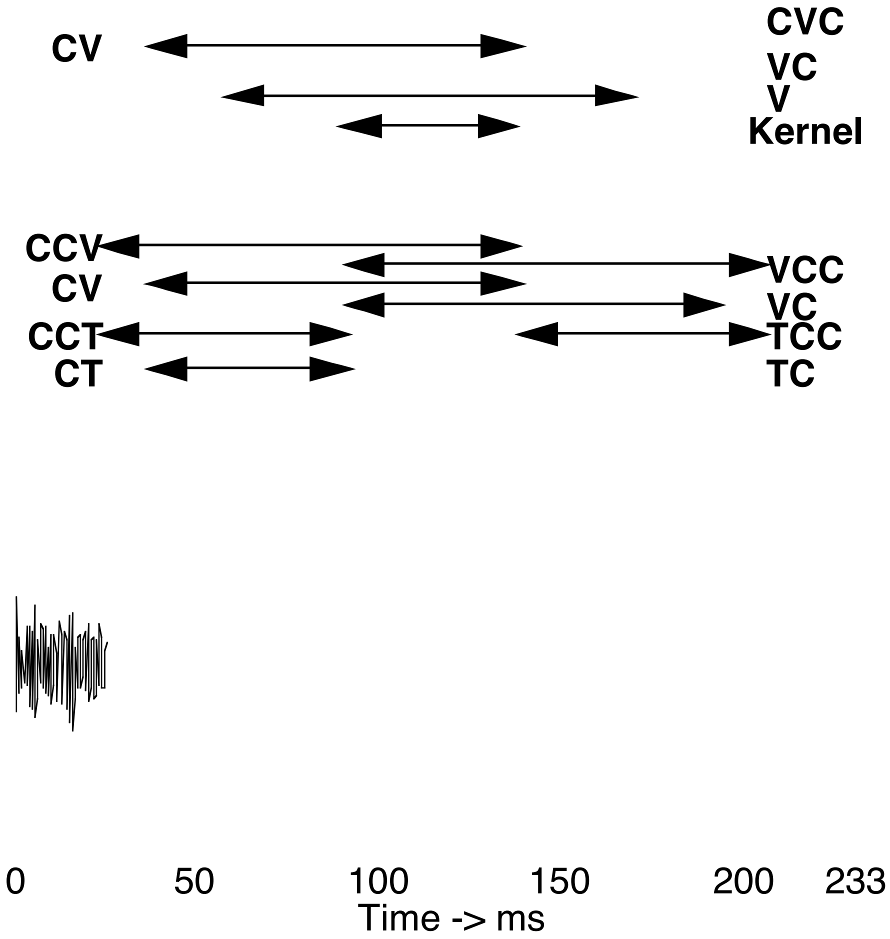
Vowel identification experiments
V*, V: Isolated synthetic vowels (two experiments)
Context: synthetic /nVf/, /fVn/ pseudosyllables
+: p<0.001 two tailed sign test

Pattern-recognition models of phoneme recognition

Strong theories (classical theories)
Presuppose strong (fixed) links between the symbolic (phoneme) and acoustic level. Strong theories of phoneme perception localize acoustic information inside the segment proper. Context information is always redundant.

Weak theories
Map cues directly to phoneme sized categories. Allow any regularity to be used for recognition. (Nearey, 1992, 1997)
Weak theories suppose that any speech can contain new (unique) information, even if it originates from the local "context".
Weak theories fit in a Pattern-recognition framework. (Smits, 1997)

Strong theories of phoneme recognition (e.g., motor theory) tend towards an obligatory phoneme hypothesis.
Weak theories of phoneme recognition tend towards a lax phoneme hypothesis


CV
CVC
VC
V
Kernel
CCV
VCC
CV
VC
CCT
TCC
CT
TC
0
50
100
150
200
233
Time -> ms