

# **Waargenomen spreektempo en articulatorische inspanning**

**Hugo Quené**

Utrechts inst voor Linguïstiek OTS, Universiteit Utrecht  
[www.hugoquene.nl](http://www.hugoquene.nl)

Dag van de Fonetiek, 20 December 2007, Utrecht

Universiteit Utrecht



# Inleiding

Luisteraars gebruiken waargenomen spreektempo om andere taalkundige verschillen te normeren...

- ▶ tussen korte–lange klinkers (**quantity**)

Nederlands *mat-maat*, Engels *beat-bead*, Japans, IJslands

[Hirata, 2004, Pind, 1996]

- ▶ tussen enkele–dubbele medeklinker

Italiaans, Japans, Fins

[Pickett et al., 1999, Hirata & Whiton, 2005, Ylinen et al., 2005]

- ▶ tussen stemloze–stemhebbende medeklinkers

verschuiving van VOT-grens, en van “best exemplar” van /b-p/, in Engels, Nederlands, Thai, Frans

[Port, 1979, Miller et al., 1984, Volaitis & Miller, 1992]

[Kessinger & Blumstein, 1997, Miller et al., 1997, Nagao & De Jong, 2007]



# Inleiding

Spreektempo wordt ook gebruikt als indicator van...

- ▶ relatieve belang, nadruk

[Zwaardemaker & Eijkman, 1928, Megehee et al., 2003]

- ▶ geslacht van spreker

[Biemans, 2000]

- ▶ vraag of mededeling

[van Heuven & van Zanten, 2005]

- ▶ spreekstijl (spontaan vs voorgelezen)

[Mixdorff & Pfitzinger, 2005]



# Circulariteit

Om spreektempo waar te nemen, moet luisteraar  
weten welke spraakklanken gehoord zijn,  
**en vice versa.**

probleem in ASR; tempo-informatie doorgaans genegeerd

[Zheng et al., 2003]

hoe slagen menselijke luisteraars  
om spreektempo te bepalen ?

tamelijk accuraat: JND 5%

[Quené, 2007]

Universiteit Utrecht



# Waargenomen spreekttempo

- ▶ aantal gesproken spraakklanken per tijdseenheid
- ▶ aantal bedoelde spraakklanken per tijdseenheid
  - semi-spontane fragmenten uit "Kiel corpus"
  - waargenomen tempo verschilt tussen Normal Sloppy en Slow Clear fragmenten, ook als beide soorten frasen zelfde tempo hebben van gesproken klanken

NS: intended /klaɪnən aʊgənblik/ ~>14.1 phone/s

realized [klaɪn aʊgəblik] ~> 9.4 phone/s

SC: intended /?unt fraita:ks/ ~> 9.4 phone/s

realized [?unt fraita:ks] ~> 9.4 phone/s

[Koreman, 2006]

- ▶ **toonhoogte:** niveau en complexiteit van bewegingen

[Kohler, 1986, Rietveld & Gussenhoven, 1987]

dit suggereert dat luisteraars rekening houden met articulatorische inspanning tijdens spraakproductie

Universiteit Utrecht



# Hypothese

Waargenomen spreektempo wordt beïnvloed door impliciete kennis (bij de luisteraars) van de articulatorische inspanning die vereist is om de gehoorde spraak te produceren.



# Algemene methode

Contrueer spraakstimuli met vast tempo,  
die alleen variëren in articulatorische inspanning.

Indien articulatorische inspanning ook perceptief  
relevant is, dan verwachten we andere  
tempo-oordelen tussen articulatorische condities.



# Stimuli 1

12 reeksen (elk 5 trocheeën, 10 lettergrepen)

- ▶ 6 met **grote** inspanning:  
bv. /tisi kuxu pyny siti xuku/  
kortere articulatorische afstanden, makkelijker
- ▶ 6 met **kleine** inspanning:  
bv. /paxi keso nuxy xipa soke/  
langere articulatorische afstanden, moeilijker

elke trochee herhaald met metronoom

"paxi paxi paxi ... paxi paxi paxi"

"keso keso ... keso keso keso keso"

goede realisaties uitgeknippt en opgelijnd op 'click track' (op klinkerinzet)

~~ "paxi keso nuxy xipa soke"

alle 12 reeksen hebben hetzelfde fysieke tempo,  
verschillen alleen in samenstellende spraakklanken



# Stimuli 2

perceptieve compensatie voor spreek-inspanning  
kan mede afhankelijk zijn van **tempo**

stimuli gemaakt voor 4 verschillende tempi:

syll/s:	3.5	4.0	4.5	5.0
bpm:	210	240	270	300
IOI:	275	250	225	200 ms



# Taak: “magnitude estimation”

- ▶ geef willekeurige waarde **100** voor tempo van eerste stimulus
- ▶ geef daarna relatieve tempo-waarden (tov 100) aan iedere volgende stimulus
- ▶ neem log van deze ruwe tempo-waarden

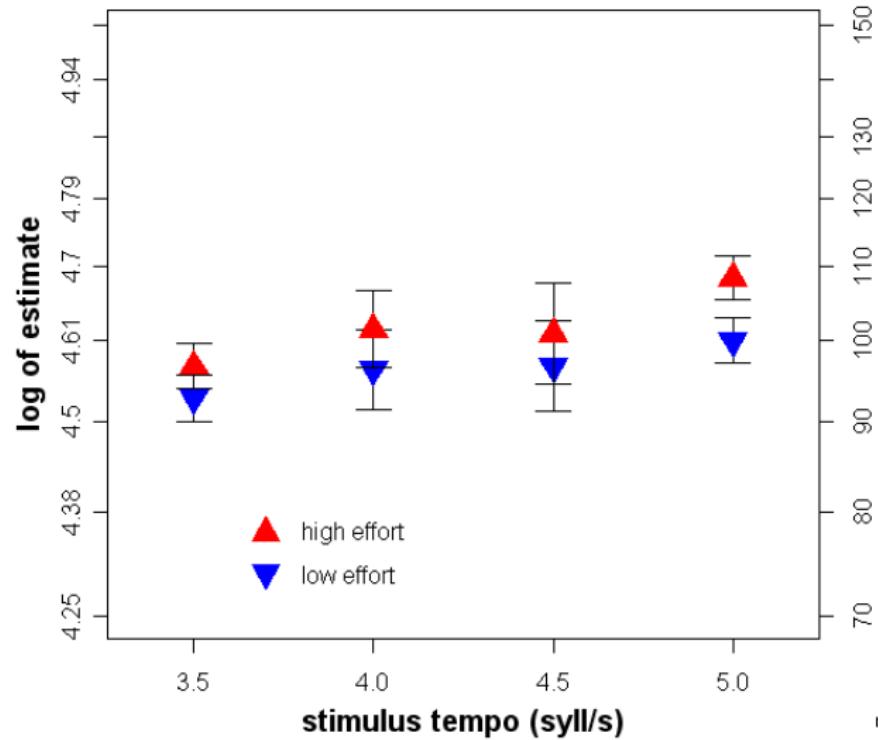
[Stevens, 1971, Grosjean & Lass, 1977]

geblokt per tempo, volgorde van blokken gevarieerd  
twee presentaties van iedere stimulus per blok  
alleen eerste responsie gebruikt in data-analyse



# Resultaten

magnitude estimation of tempo



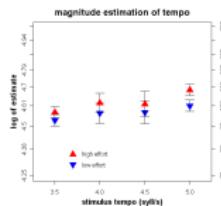
# Mixed-effects modellering

- Gecombineerd model met twee 'fixed' effecten (4 tempi, 2 condities) en twee gekruiste 'random' effecten (34 ppn, 12 items)

[Bates, 2005, Quené & van den Bergh, 2004,

Quené & van den Bergh, 2007]

	mean	s.e.	
fixed intercept	4.525	(0.044)	
$\gamma_{240}$	0.035	(0.020)	
$\gamma_{270}$	0.020	(0.020)	
$\gamma_{300}$	0.101	(0.020)	$p < .001$
$\gamma_H$	0.059	(0.019)	$p = .002$
random	mean	95% C.I.	
stimuli $\sigma_{v_0(0k)}^2$	0.0005	n/a	
listeners $\sigma_{u_0(j0)}^2$	0.053	(0.034, 0.095)	
residual $\sigma_{e_i(jk)}^2$	0.078	(0.073, 0.084)	
evaluation			
-2 log(lhr)	587.8		



# Discussie

- ▶ articulatorische inspanning heeft inderdaad invloed op waargenomen (subjectieve) spreektempo



# Discussie

deze wisselwerking tussen productie en perceptie is consistent met

- Motor Theory of Speech Perception
- compensatie voor coarticulatie  
(articulatorische gebaren)

luisteraars houden rekening met fysieke beperkingen  
(embodiment) van spreker



# Conclusie

articulatorische inspanning tijdens spraakproductie heeft invloed op waargenomen spreektempo tijdens spraakperceptie



# V R A G E N ?





# Stimulus sequences

**Table:** Stimulus sequences used in the magnitude estimation experiment.

nr	effort	sequence
B	mixed	/paxi siti keso xipa tisi/
1	low	/tisi kuxu pyny siti xuku/
2	low	/kuxu pyny tisi xuku nypy/
3	low	/pyny tisi kuxu nypy siti/
4	low	/siti xuku nypy tisi kuxu/
5	low	/xuku nypy siti kuxu pyny/
6	low	/nypy siti xuku pyny tisi/
7	high	/keso nuxy paxi soke xynu/
8	high	/nuxy paxi keso xynu xipa/
9	high	/paxi keso nuxy xipa soke/
10	high	/soke xynu xipa keso nuxy/
11	high	/xynu xipa soke nuxy paxi/
12	high	/xipa soke xynu paxi keso/



# References 1

-  Bates, D. (2005). Fitting linear models in R: Using the `lme4` package. *R News*, 5(1):27–30.
-  Biemans, M. (2000). *Gender variation in voice quality*. PhD Thesis, Katholieke Universiteit Nijmegen.
-  Some factors affecting the listener's perception of reading rate in English and French. *Language and Speech*, 20(3):198–208.
-  Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *J. Phonetics*, 32(4):565–589.
-  Hirata, Y. and Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanese. *J. Acoust. Soc. Am.*, 118(3):1647–1660.
-  Kessinger, R. H. and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *J. Phonetics*, 25(2):143–168.
-  Kohler, K.J. (1986). Parameters of speech rate perception in German words and sentences: Duration, F0 movement, and F0 level. *Language and Speech*, 29(2):115–139.
-  Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *J. Acoust. Soc. Am.*, 119(1):582–596.
-  Megehee, C. M., Dobie, K., and Grant, J. (2003). Time versus pause manipulation in communications directed to the young adult population: Does it matter? *J. Advertising Research*, 43(3):281–292.
-  Miller, J. L., Aibel, I., and Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perc. and Psychoph.*, 35(1):5–15.



# References 2

-  Miller, J. L., O'Rourke, T. B., and Volaitis, L. E. (1997). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, 54(3–4):121–137.
-  Mixdorff, H. and Pfitzinger, H. R. (2005). Analysing fundamental frequency contours and local speech rate in map task dialogs. *Speech Commun.*, 46(3–4):310–325.
-  Nagao, K. and De Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *J. Acoust. Soc. Am.*, 121(5):2882–2898.
-  Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, 56(3–4):135–157.
-  Pind, J. (1996). Rate-dependent perception of quantity in released and unreleased syllables in Icelandic. *Speech Commun.*, 19(4):295–306.
-  Port, R. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *J. Phonetics*, 7:45–56.
-  Quené, H. (2007). On the just noticeable difference for tempo in speech. *J. Phonetics*, 35(3):353–362.
-  Quené, H. and van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Commun.*, 43(1–2):103–121.
-  Quené, H. and van den Bergh, H. (submitted). Examples of mixed-effects modeling with crossed random effects and with binomial data.
-  Rietveld, A. and Gussenhoven, C. (1987). Perceived speech rate and intonation. *J. Phonetics*, 15:273–291.



# References 3

-  Stevens, S.S. (1971). Issues in psychophysical measurement. *Psychological Review*, 78:426–450.
-  van Heuven, V. J. and van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Commun.*, 47(1–2):87–99.
-  Volaitis, L. and Miller, J. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.*, 92(2):723–735.
-  Ylinen, S., Shestakova, A., Alku, P., and Huotilainen, M. (2005). The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. *Language and Speech*, 48(3):313–338.
-  Zheng, J., Franco, H., and Stolcke, A. (2003). Modeling word-level rate-of-speech variation in large vocabulary conversational speech recognition. *Speech Commun.*, 41(2–3):273–285.
-  Zwaardemaker, H. and Eijkman, L.P.H. (1928). *Leerboek der Phonetiek: inzonderheid met betrekking tot het Standaard-Nederlandsch*. Haarlem: Erven F. Bohn.

