# When do listeners think about the speaker?

## Neural evidence for rapid voice-based speaker modelling

Jos van Berkum[1,2] Dannie van den Brink[2,3] Cathelijne Tesink[3,2] Miriam Kos[2] & Peter Hagoort[2,4]

1 Univ of Amsterdam, Psychology 2 FC Donders Centre for Cognitive Neuroimaging 3 Radboud Univ, Psychiatry 4 Nijmegen Institute for Cognition and Information

## The issue

We all use knowledge about other people to make sense of what they say. For instance, we know that "Next year I will quit smoking" is an unlikely thing to say for a four-year old, and that "I just visited my gynecologist" is really odd for a man to say. Examples such as these reveal that at some point during language comprehension, people relate what's being told to who is telling it. We used event-related brain potentials to determine exactly when and how listeners do this.

## The experiment

We recorded the EEG from 12 male and 12 female Dutch listeners as they listened to a mix of sentences delivered by 21 different speakers. Participants, who weren't given any additional task, heard 80 utterances whose lexical content did not easily fit probabilistic inferences about the speaker's sex, age, and social-economic status, inferred from the speaker's voice. The speaker inconsistency always emerged at a single critical word. Although some speaker inconsistencies were truly anomalous, most of them merely violated social stereotypes (see panel A for examples; critical word is in bold, inconsistent speaker in red).

The 80 speaker-inconsistent utterances were mixed with 80 speaker-consistent control utterances and 192 additional utterances, of which 48 contained a classic sentence-dependent lexical-semantic anomaly, known to elicit a large N400 effect relative to a correct control (see panel B for examples; semantic anomaly is in red).

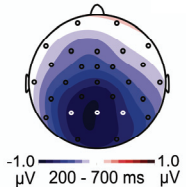
Partly funded by an *NWO Innovation Impulse* grant to JvB. Thanks to Oliver Müller, Petra van Alphen, & José Kerstholt for help. Poster file downloadable from www.josvanberkum.nl. (berkum@psy.uva.nl). CNS-2006, San Francisco Apr 8-12

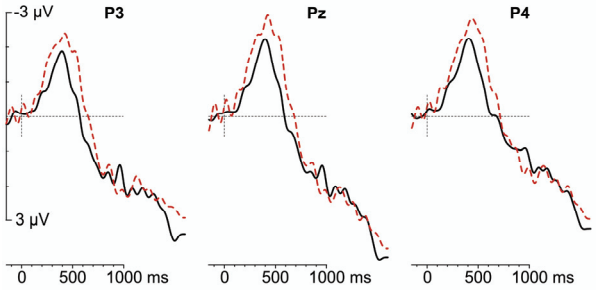
### Speaker inconsistency effect — A

*male / female :* "If only I looked like **Britney** Spears in her latest video"

*upper- / lower-class :* "I have a large **tattoo** on my back"

*young child / adult :* "Every evening I drink some **wine** before I go to sleep"



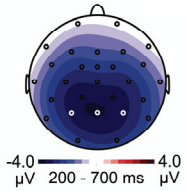



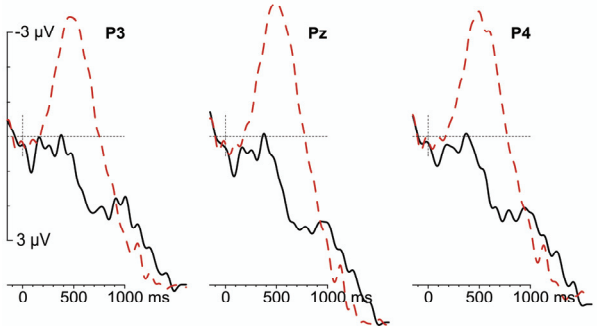

### Semantic anomaly effect — B

"Dutch trains are **sour** and blue"
"Dutch trains are **yellow** and blue"

"You wash your hands with **horse** and water"
"You wash your hands with **soap** and water"

"The earth revolves around the **trouble** in a year"
"The earth revolves around the **sun** in a year"




## Results

Speaker-dependent inconsistencies elicited a small but reliable posterior ERP effect from about 200-300 ms onwards (panel A). Although smaller, this effect is qualitatively identical to the standard N400 effect elicited by semantic anomalies (panel B)[1]. This suggests that voice-inferred information about the speaker's identity is taken into account by the same early language interpretation mechanisms that construct 'sentence-internal' meaning based on just the words.

## Implications

Our findings directly challenge the classic assumption that the meaning of a sentence is initially composed out of the fixed meaning of individual words, independently of the wider context in which it is used. Instead, listeners use what they know or can infer about the speaker in the earliest stages of meaning construction. As such, our findings testify to the fact that language comprehension is immediately contextualized and that linguistic meaning cannot be separated from the social context of language use. The latter makes sense, as language and language competence evolved in face-to-face social interaction.

What the brain's electrophysiology allows us to see is that information encoded in the meaning of spoken words immediately merges with voice-based cues about the identity of the speaker, and that it takes only a few hundred milliseconds to relate what's being told to who's telling it. Linguistic meaning is deeply social.

[1] The N400 effect size difference reflects, amongst other things, the larger average severity of the lexical anomalies, and the fact that voice-based cues to the speaker's identity are inherently probabilistic (e.g. a woman can have a low voice).