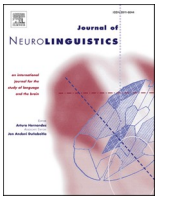
Journal of Neurolinguistics 74 (2025) 101248

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling



Does sounding 'Gay' or 'Straight' affect how we understand language? Sentence comprehension is regulated by the speaker's perceived sexual orientation

Simone Sulpizio [a,b,*], Fabio Fasoli [c,d], Gaia Lapomarda [e], Francesco Vespignani [f,g,h]

[a] Department of Psychology, University of Milano-Bicocca, Milan, Italy
[b] Milan Center for Neuroscience (NeuroMI), University of Milano-Bicocca, Milan, Italy
[c] School of Psychology, University of Surrey, Guildford, United Kingdom
[d] Instituto Universitário de Lisboa (ISCTE-IUL), Portugal
[e] Institute for Psychology, University of Innsbruck, Austria
[f] Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy
[g] Intredipartimental Research Center "I-APPROVE–International Auditory Processing Project in Venice", University of Padua, Padua, Italy
[h] Padova Neuroscience Center (PNC), University of Padua, Padua, Italy

ARTICLE INFO

Keywords:
Sentence comprehension
ERPs
Social identity
Sexual orientation
Voice
Sounding gay

ABSTRACT

Social interactions are shaped by the way individuals communicate. Listeners form impressions based on how someone sounds, and the message conveyed can be interpreted differently depending on who the speaker is. We investigated on-line sentence processing focusing on the role of the speaker's gay- vs. heterosexual-sounding voice in the construction of meaning. Event-related brain potentials were recorded while participants listened to two gay- and two heterosexual-sounding male speakers uttering stereotypical sentences. We manipulated whether the sentences referred to professions stereotypically congruent or incongruent with the speakers' perceived sexual orientation. Results showed that the interplay between the speaker's voice and message content influenced sentence processing early after an incongruent stereotype was presented. The interaction was maximal at frontal sites, with a larger negativity for stereotypically-congruent than for stereotypically-incongruent professions when uttered by gay-sounding speakers. These results suggest that the perception of the speaker as gay- or straight-sounding is quickly used by listeners to build the message meaning. The inconsistency between vocal and linguistic information modulates a frontal negativity, potentially indicating control processes during sentence comprehension put in place to deal with the inconsistency.

1. Introduction

Everyday verbal communication requires listening and interpreting the meaning of the messages that have been conveyed. To understand what the speaker means, the listener needs to consider what has been said (the message) but also who said it (the speaker). Indeed, the literal meaning of the message can be interpreted differently depending on how the speaker is perceived. Vocal cues are used by listeners to categorize speakers in given social categories (see Belin et al., 2004) and develop stereotypical expectations about

* Corresponding author. Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milano, Italy.
E-mail address: simone.sulpizio@unimib.it (S. Sulpizio).

https://doi.org/10.1016/j.jneuroling.2025.101248
Received 6 June 2024; Received in revised form 12 February 2025; Accepted 17 February 2025
Available online 22 February 2025

the speaker and what they may say. For instance, individuals may be surprised to hear a male speaker saying '*I like wearing make-up*' because they would be more likely to expect a woman rather than a man to engage in such stereotypically feminine behavior. Understanding how stereotype-based messages are processed is extremely important as this can contribute to understand how such stereotypes are transmitted and perpetuated over time (see Lyons et al., 2007; Lyons & Kashima, 2003), and also how individuals interact socially. To fully understand this, we need to examine how and when the message and the speaker's voice are integrated. More specifically, is the information inferred from the voice immediately used to actively build meaning? Or does it become relevant only at a later stage of processing? This is particularly important when the voice conveys information concerning the speakers' minority status and the message is interpreted accordingly. We focused here on sexual orientation, since gay men are often stereotyped, perpetuating their stigmatization and relegation to specific social roles (see Fasoli et al., 2017). Understanding how listeners process and integrate vocal cues in processing message content will shed light on how stereotypes are maintained via communication.

In this work, we focused on brain activity and used event-related brain potentials (ERPs) to track the time dynamic of speaker-message integration, that is when listeners integrate the message content and the social information from the speaker's voice. We focused on male voices that were perceived as either gender 'typical' or 'atypical' and, hence, straight- or gay-sounding, respectively. We examined if, how, and when sounding gay or straight affects online mechanisms of sentence comprehension of messages that involved gender stereotypes (i.e., professions), namely expectations that are overgeneralized to all members of each group regardless of whether they are true or not (Ellemers, 2018). Indeed, gender stereotypes are attributed differently to gay- and straight-sounding male speakers (Fasoli et al., 2017) and play a role in message interpretation (Fasoli et al., 2020).

1.1. The speaker-meaning integration in the building of sentence meaning

Although at the core of linguistic communication, the issue of speaker-meaning integration, and more generally how information about the speaker's perceived social information is integrated with the explicit linguistic information, has been largely overlooked in the cognitive neuroscience of language. So far, models of sentence comprehension have been mainly developed in the mainstream tradition of cold cognition (Jacobs, 2015), following seminal modular views of cognition. In this tradition, the basic compositional linguistic processes have been studied in isolation from aspects of everyday communicative situations, including (a) the fact that message and voice are unlikely to be processed by separate modules with a limited possibility of interaction and (b) that some inferential processes can depend on the listeners' beliefs and attitudes. ERP research on gender stereotypes did not consider these aspects (Osterhout et al., 1997) and limited evidence exists on how the ERP response is modulated by individuals' attitudes and beliefs (e.g., Canal et al., 2015; Jiang et al., 2020).

Remarkable research on speaker-message integration is represented by the seminal study by van Berkum and colleagues (Van Berkum et al., 2008). They measured ERPs while participants listened to sentences uttered by speakers whose voices were either stereotypically consistent or inconsistent with the message content (e.g., the sentence "*Every evening I drink some wine before I go to sleep*" uttered by either an adult or a child's voice). The speaker-meaning inconsistency was tested in relation to the speaker's gender (man/woman), age (young/old), or socioeconomic status (low-/high-class). Results showed that the speaker-meaning inconsistency elicited an N400 effect identical (in terms of timing and scalp distribution) to the classic effect elicited by a semantic anomaly (e.g., sour in "*Dutch trains are sour and blue*") but much smaller in amplitude. The N400 is an index of lexical-semantic processing and reflects the easiness of lexical and conceptual retrieval and integration (Lau et al., 2008). In sentence processing, a larger N400 is associated with words that are semantically inconsistent with or highly unexpected from the context. Recent evidence suggests that the N400 modulation is affected by predictability, which likely impacts lexical and conceptual access (e.g., Kuperberg et al., 2020; Lau et al., 2016). van Berkum et al. (2008) concluded that "voice-inferred information about the speaker is taken into account by the same early language interpretation mechanisms that construct "sentence-internal'' meaning based on just the words" (p. 586). This finding is in line with the multiple constraints approaches to sentence comprehension in which any kind of information, including pragmatic and extralinguistic information, is processed as soon as available and simultaneously integrated by a unique compositional mechanism (e.g., Unification; Hagoort et al., 2004). The fact that both the timing (when) and the topography (how) of the ERP effect are very similar to lexical-semantic violations is important within this proposal. Indeed, Hagoort et al. (2009) suggested that "world knowledge, information about the speaker, co-occurring visual input, and discourse information all trigger electrophysiological responses similar to those triggered by sentence-internal semantic information" (p. 832).

Contrasting results, however, have been reported by Lattner and Friederici (2003), who asked participants to listen to sentences which were either stereotypically gender congruent (a man uttering "*I like to play soccer*") or incongruent (a man uttering "*I like to wear lipstick*") with the speaker's gender. They found that the voice-message incongruence elicited a posterior positivity, interpreted as a P600, indexing a stage of re-processing in which the listener integrates the message meaning with the stereotypical inferences about the speaker. Also, Foucart et al. (2015) adapted van Berkum et al.'s (2008) paradigm to examine how Spanish bilinguals process speaker-message inconsistencies related to gender and age. This study found no N400 (even in the L1 group) but a posterior positivity (dubbed Late Positive Potential), which had an earlier onset in L2 speakers. This result was interpreted in line with Lattner and Friederici (2003). Foucart et al. (2015) account for the failure to replicate van Berkum et al.'s (2008) result by considering the possibility of unbalance in cloze-probability of the van Berkum et al.'s (2008) material, a potential overlap of late positivities with the N400 and, finally, individual differences. Thus, in contrast with van Berkum et al. (2008), these results suggest that the integration of the speaker's voice and message content during sentence processing may take place at a rather later stage and may affect processing mechanisms that are distinct from those dealing with compositional word level processing (namely Unification). In other words, according to Lattner and Friederici (2003), the speaker's social information conveyed by voice does not affect the meaning of the words, but rather how difficult it is to reconcile the message with listeners' expectations about the speaker. This may be particularly

relevant when speakers' voices convey information about 'ambiguous' social categories.

1.2. Vocal cues of sexual orientation and gender typicality

Sexual orientation is an 'ambiguous' social category that cannot be ascertained until the speaker self-discloses their sexuality (Tskhay & Rule, 2013). Still, voice is considered a sexual orientation cue (Fasoli et al., 2016). Listeners consistently categorize speakers' sexual orientation according to how they sound, even after listening to words or single syllables (Gaudio, 1994; Munson, 2007; Sulpizio et al., 2015, 2020). This categorization process is not always accurate and often occurs based on stereotypical ideas of how gay men sound (Kachel, Simpson, & Steffens, 2018; Smyth et al., 2003; Sulpizio et al., 2015). A gay-sounding speaker is often perceived as deviating from 'normative' ways of speaking that, for a man, involves sounding masculine (Munson, 2007; Zimman, 2013) and not engaging in stereotypical 'gay' speech (e.g., lisping; Mack & Munson, 2012). Indeed, speakers are assumed to be straight – a phenomenon called 'straight categorization bias' (Lick & Johnson, 2016) – unless they sound gender 'atypical' (Fasoli et al., 2023; Masi and Fasoli, 2022) and men are aware of the relationship between perceived gender typicality and assumed sexual orientation (see Fasoli et al., 2018).

In line with the gender inversion theory (Kite & Deaux, 1987), gay men are often perceived as lacking stereotypically masculine traits and possessing more feminine characteristics and interests (i.e., professions and activities; Blashill & Powlishta, 2009). Such stereotypical expectations on personality traits, interests/hobbies, and professions emerge even when individuals listen to gay- and straight-sounding male speakers (see Fasoli & Maass, 2018). Moreover, the lack of attribution of stereotypically masculine traits explains whether gay-sounding male speakers are not seen as a good fit for stereotypically masculine roles (Fasoli et al., 2017). Hence, not only voice triggers sexual orientation categorization, but also expectations on social roles, including professions. So far, only a few studies have examined the interplay between vocal cues of sexual orientation and message content. Recently, Fasoli & Formanowicz, 2024 showed that gay-sounding male speakers uttering agentic messages can counteract the stereotype that associates gay-sounding men with a lack of agency. Importantly for this study, ambiguous messages uttered by gay-sounding and straight-sounding speakers were interpreted according to stereotypical gender expectations. For instance, a male speaker stating that he played sports was more likely thought to be assumed to play football if he sounded straight than gay (Fasoli et al., 2020). While these findings are promising, they are silent on the how voice-based speaker information and message content are processed online and interact.

1.3. The present study

Although all previous studies suggest that listeners extract information about the speakers' social information from their voices and use it during sentence processing, they are inconclusive about the two main issues. First, *when* this happens. Second, *whether* the system integrates the word meaning and the information extracted from voice by a unique mechanism – as suggested by van Berkum et al. (2008) – or by taking the latter into account only in a second stage – as suggested by Lattner and Friederici (2003) and Foucart et al. (2015).

To address these issues, we ran an ERP experiment in which participants listened to sentences spoken by speakers whose voices were either stereotypically congruent or incongruent with the sentence content (as in van Berkum et al.'s paradigm). In particular, sentences were all linguistically acceptable but contained one word at which the linguistic message started to match or mismatch with the stereotypical beliefs about the speaker's social information extracted from the voice. Extending previous work, we focused on speakers who sounded either gay or straight. Previous studies focused on social categories assumed from voice referring to gender and age (Lattner & Friederici, 2003; Foucart et al., 2015; Van Berkum et al., 2008), which are less ambiguous than sexual orientation. Although sexuality is fluid (Diamond, 2016) and multiple sexual orientations exist (e.g., bisexuality, pansexuality), we only considered male speakers who identified as either gay or straight and who were likely to be perceived as such because of the way they sounded. We also only considered cisgender male speakers (i.e., individuals who identified as men and were assigned male at birth). This allowed us to consider speakers who were similar in terms of biological features (e.g., larynx morphology and sex-related hormones; e.g., Dabbs Jr & Mallinger, 1999) but varied in perceived sexual orientation.

Neuroimaging studies indicate that perception or knowledge of social categories and objects recruits different cortical networks (e.g., Contreras et al., 2012). Previous studies on speaker-message integration may have mixed manipulations of different natures. These studies used or included gender as a social category. Gender is a socially relevant category, for which most cultures show strong stereotypes. However, it may not be the best test case to study the interaction between conceptualization linked to social categories and language, as stereotypical gender interacts with grammatical gender in different ways across languages. ERP evidence indicates that the neural processing of gender retrieval differs for words bearing *definitional* gender (e.g., actress, father) and words bearing *stereotypical* gender (e.g., teacher, driver; e.g., Siyanova-Chanturia et al., 2012). Hence, there is a need to consider other social categories and focus on perceived rather than actual social categories. Moreover, a scrutiny of the examples of the material used by Van Berkum et al. (2008) shows that, even if all the examples are instances of a strong inconsistency between the speaker's perceived social category and the message, the type of knowledge that triggers the inconsistency was rather inhomogeneous. Target words can be more or less frequently used by a given category of speakers (e.g., wine vs milk, said by a speaker perceived to be a child) and this feature, and not the incongruity with the message, could modulate the amplitude of the N400. Moreover, concerning gender, when thinking about cisgender men and women, some stimuli could have been perceived as more likely to be said in real life than others. For instance, sentences referring to bodily characteristics (e.g., a cisgender man saying "*I recently had a check-up at the gynecologist in the hospital*" or "*I might be pregnant because I feel sick*") may be perceived as particularly unlikely compared to other sentences that merely violate stereotypical expectations (e.g. a man saying "*When I watch TV I often cry during a good movie*"). These observations suggest that the

modulation of the N400 found by van Berkum et al. (2008) could occur not because these violations are perceived as weaker than sentence-internal ones ("*Dutch trains are sour and blue*"), but rather because a (little) part of the material elicited an N400 at the target word that is linked to different phenomena (e.g., speaker-dependent frequency of usage or type of content). To overcome some of these limitations, we used stimuli, namely professions, that were consistent in terms of stereotype violation. Sentences were all linguistically acceptable and contained one target word (i.e., profession) at which the linguistic message matched or mismatched with the stereotypical beliefs about the speaker's social information extracted from the voice. Stereotypically masculine and feminine professions are differently associated with gay and straight men (Blashill & Powlishta, 2009), even when sexual orientation is inferred from voice (Fasoli & Maass, 2018). Hence, we manipulated message stereotypicality by using sentences in which the speaker referred to himself as engaging in a stereotypically masculine (e.g., "*My friends always tell me that I'm a very good footballer to convince me to donate them some tickets*") or stereotypically feminine profession (e.g., "*My friends always tell me that I'm a very good hairdresser to convince me to donate them a haircut*"). Therefore, a gay-sounding male speaker saying that he is involved in a stereotypically feminine job represented a stereotype-congruent situation, while the same speaker referring to a stereotypically masculine job indicated a stereotype-incongruent situation.

Finally, focusing on perceived sexual orientation allowed us to investigate if the process of detection of speaker-message inconsistency is affected by listeners' individual differences (as suggested by Foucart et al., 2015). Attitudes toward and familiarity with gay men could contribute to modulating the relevance of the inferred speaker's social category (i.e., perceived sexual orientation) in message comprehension. While everyone has experience and contact with 'marked' social categories (e.g., gender and age), a different extent of contact with members of 'ambiguous' social categories like sexual orientation may occur, as this assumes gay people to have disclosed their sexuality. This makes familiarity with gay men an interesting individual difference to consider. Indeed, familiarity with gay men can affect expectations and stereotyping (Herek, 1996; Salvati et al., 2019). Moreover, negative attitudes toward gay men are associated with a stronger perception of them as violating gender role expectations (Lehavot & Lambert, 2007). Hence, we tested the associations between message interpretation and listeners' familiarity and attitudes toward gay men.

1.4. Hypotheses

In terms of expected ERPs modulation, we advanced the following predictions for the two groups of speakers. For straight-sounding male speakers, if information about the speaker's social category is inferred from voice quickly and efficiently, and is integrated as soon as it becomes available with the linguistic message content using the same neuro-cognitive mechanisms that compute meaning from words, then the speaker-message inconsistent condition (e.g., "*My friends always tell me that I'm a very good hairdresser to convince me to donate them a ticket for a free haircut*") should elicit a larger N400 at the underlined words than the speaker-message consistent condition (e.g., "*My friends always tell me that I'm a very good footballer to convince me to donate them a ticket for the match*"). This result would be in line with van Berkum et al.'s (2008) findings. However, if the speaker's social category inferred from voice is taken into account only later on by the linguistic system, then the speaker-message inconsistency should be associated with a late posterior positivity. This result would be in line with Lattner and Friederici's (2003) findings. Concerning the gay-sounding speakers, two different predictions may be advanced: The effect might reverse compared to straight-sounding speakers, showing larger N400/P600 for stereotypically masculine than feminine professions; in contrast, the effect elicited by gay-sounding speakers might simply be smaller than that elicited by heterosexual-sounding speakers. In both cases, a significant interaction between the speaker's social category and profession is expected.

Finally, we explored the relationships between such effects and listeners' individual differences meaning their levels of sexual prejudice and familiarity with gay men.

2. Method

2.1. Participants

Forty-three cisgender Italian participants took part in the experiment. Seven participants were excluded from the analyses because of the very high noise in the ERP recording and the massive presence of muscular artifacts. Since lesbian, gay, and bisexual individuals show differences in the way they categorize speakers' sexual orientation (i.e., they use the 'gay label' in a less strict way than heterosexual individuals; Fasoli et al., 2021) and show more positive attitudes toward gay men (Jellison et al., 2004) than heterosexual, we excluded one participant who self-identified as bisexual as they may form a different stereotypical impression about the speaker. The final sample consisted of 35 cisgender heterosexual participants (26 women, mean age: 24.22, sd: 3.87). All participants were native Italian speakers, right-handed, with normal or corrected-to-normal vision, and neurologically healthy.

2.2. Materials

2.2.1. Sentences

We created 216 sentences involving stereotypically masculine and feminine professions (for the full list of stimuli see Supplementary Materials). Thirty-six stereotypically masculine professions and 36 stereotypically feminine professions were selected from Misersky et al.'s database (2014). Forty-six professions were grammatically marked as gender neutral, whereas for the remaining (14 of the feminine and 12 of the masculine professions) grammatical gender was explicitly expressed by the inflection – in this case, professions were always declined as masculine and always agreed in gender with the other words of the noun phrase (no female epicene

was used). The two sets of professions were matched on: word frequency (extracted from SUBTLEX-IT, Crepaldi et al., 2013), number of letters, and number of phonemes (all *ps* > .15). To vary the association between the critical word and the sentence context, each word was associated with three different contexts. Thus, the professions were included in 108 host sentences that had no bias for gender or SO; each sentence context was associated with two professions so that once it contained a stereotypically masculine profession (e.g., *calciatore*, footballer) and once a stereotypically feminine profession (e.g., *parrucchiere*, hairdresser). We built 12 lists so that each profession was presented in all three sentence contexts and uttered by all the speakers, but each profession was presented only once to each participant. Participants listened to sentences in random order.

Sentences were created so that the stereotypical consistency/inconsistency between speaker's voice and stereotypical content always emerged at a single critical word, and this word was never the first – to give listeners vocal cues to make assumptions about the speaker's sexual orientation – or the last of the sentence – to avoid that our effects would depend on sentence-final wrap-up processes. Except for these two positions, the position of critical words varied across sentences. Gender disagreement between the target noun and the other words of the phrase never occurred, and stereotypicality of the target was not predictable from pre-target words.

An example of a sentence is "*Dopo anni di studio sono diventato ballerino della Scala*" ("*After years of studying, I'm now a dancer at the Scala theater*"). The sentence contained a stereotypically-feminine profession (i.e., dancer) and was thus assumed to be perceived as stereotypically consistent when uttered by a gay-sounding rather than a straight-sounding speaker (for a full list of sentences see Supplementary Materials).

2.2.2. Speakers

All sentences were recorded by 4 cisgender male speakers. They were from the same region (Veneto), were highly educated (they held either an MSc or PhD), and were all young adults (age range: 30–32). Two speakers self-identified as gay and the other two self-identified as straight. The acoustic duration of target professions did not differ between self-identified gay and straight speakers (M_{gay} = .49 s, *SD* = .09, $M_{straight}$ = .60 s, *SD* = .05 $p > .5$). A pretest (N = 15, 10 females) was conducted to test perception of speakers while pronouncing 3 sentences (one involving a stereotypically feminine job, one referring to stereotypically masculine job and one a stereotypically gender-neutral job). Each speaker uttered a sentence at the time (randomly presented) and listeners rated the speaker on the following dimensions. First, they judged the speakers' sexual orientation on a 7-point Likert scale from 1(*exclusively heterosexual*) to 7(*exclusively gay*). Then, participants indicated how masculine, feminine, and likeable the speakers were on a 7-point Likert scale from 1(*not at all*) to 7(*completely*). We included likeability ratings to ensure the speakers were perceived as similarly likeable and to avoid a positive bias toward certain speakers. No significant difference emerged in terms of perceived sexual orientation, pairwise *t*-test: $t(14) = -2.00$, $p = .06$, but the mean for the straight speakers ($M = 2.73$, $SD = 1.05$) were descriptively lower than that of gay speakers ($M = 3.60$, $SD = 1.29$) in line with previous work on auditory gaydar (cf. Sulpizio et al., 2015 for similar means when judging sexual orientation from voice on a Likert scale; this is also indicative of a straight categorization bias consistently found in auditory gaydar research, see Fasoli et al., 2023). Also, straight speakers were perceived as more masculine ($M = 4.83$, $SD = 1.24$) than gay speakers ($M = 2.40$, $SD = .98$), $t(14) = -4.67$, $p < .001$, while no difference was found in femininity and likeableness (*ts* < .90, *ps* > .38). In line with auditory gaydar work, we found that gay speakers as lacking masculinity and tended to be perceived as relatively different from straight speakers in terms of sexual orientation (see Fasoli et al., 2023; Munson, 2007). Thus, we refer here to the two groups of speakers as *gay-sounding* and *straight-sounding* speakers.

2.2.3. Sexual prejudice and familiarity

We were interested in blatant hostile sexual prejudice that is related to behavioral and affective attitudes (see Preuss et al., 2020) and discriminatory intentions (Huic et al., 2018). We assessed prejudice by asking participants to complete the Attitudes Towards Gay Men scale (Herek & McLemore, 1998), in which they had to indicate their agreement with 10 items (e.g., "I think male homosexuals are disgusting") on a 7-point Likert scale from 1 (*completely disagree*) to 7 (*completely agree*). This scale has been widely used and has demonstrated strong reliability in measuring individuals' biases and attitudes (α = .92 in this study), consistently producing reliable results across various populations and settings (Grey et al., 2013). Ratings were averaged to create a score. The higher the score, the stronger the prejudice toward gay men.

Participants were also asked to report their level of familiarity with gay men by answering 5 items (e.g., "How often do you interact with openly gay men?"; Brambilla et al., 2013) on a scale from 1 (*never*) to 7 (*always*). The scale has been used in the context of gaydar studies and showed good internal reliability in our study (α = .75). Ratings were averaged to calculate a score, with a higher score indicating greater familiarity with gay men. Typically, individuals who are more familiar with gay men are more accurate in judging men's sexual orientation (Brambilla et al., 2013).

2.3. Procedure

At their arrival, participants were fully informed about the nature of the study and provided their written informed consent to participate. Then, we montaged the electrode cap. After that, participants sat in a sound-attenuated, electrically shielded boot, in front of a computer screen and two speakers were used for stimuli delivery. Each trial started with a fixation cross presented in the center of the screen; when participants were ready they pressed the spacebar to proceed with the trial; the trial sequence was as follows: a blank of 100 ms, a short presentation of the fixation cross centered on the screen (100 ms), a blank of 100 ms, the fixation cross and the play of the stimulus, with the fixation remaining on the screen until the end of the sentence. Participants were exposed to sentences involving stereotypically masculine and feminine professions uttered by gay-sounding and heterosexual-sounding male speakers. After a blank of 500 ms, a comprehension question referring to the sentence content (e.g., *"Does the speaker sell medicines?"*) was asked in 8%

of the trials and participants answered the question with a yes/no response provided by means of labeled keyboard buttons (m/c). Comprehension questions were introduced to make sure that participants were paying attention to the message content considering that they were exposed to several sentences, without (otherwise) doing nothing. Each participant heard the sentences involving all the 72 target professions, 18 for each speaker. A brief practice preceded the experiment. The experiment was run using E-Prime software (version 2.0, Psychology Software Tools, Pittsburgh, PA; www.pstnet.com).

After the EEG session, participants completed a short questionnaire in which they reported their demographic information (age, gender, sexual orientation) and indicated how many of the speakers they had listened to during the experiment they believed were gay (from 0 to 4). Finally, we assessed sexual prejudice and familiarity with gay men.

The study received ethical approval from the University of [masked] Ethics Committee (protocol number: 2013-006).

2.4. EEG recording and data processing

EEG was recorded from 64 scalp electrodes (Fp1, Fpz, Fp2, AF7, AF3 AF4, AF8, F9, F7, F5, F3, F1, Fz, F2, F4, F6, F8, F10, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO4, PO8, O1, Oz, O2) positioned on an elastic cap according to the international standard 10%. An additional external electrode was placed below the left eye. All sites were referenced to the left mastoid and the ground was placed at AFz. Impedance was kept below 10 kΩ. Data were acquired at the sampling rate of 250 Hz with a low-pass filter with 100 Hz cutoff frequency and a high-pass filter with a 10s time constant. Offline analysis was performed with Brain Vision Analyzer II software (Brain Products GmbH, Munich, Germany). The EEG was filtered with a 40 Hz low-pass filter (12 dB/oct) and a high-pass filter (3.1 s time constant 12 dB/oct) after acquisition. The continuous signal was then corrected for eye blinks and ocular movements by using Independent Component Analysis and re-referenced to the linked mastoids. The EEG was segmented between 200ms before the onset of the critical word and 1000ms after. Epochs were visually inspected to reject those contaminated by artifacts. In this way, 6.2% of the trials were excluded before ERP average. Single participant waveforms for each condition were averaged after correcting the baseline to the mean of the 150 ms pre-target interval.

3. Results

3.1. Preliminary analyses

Overall participants reported low levels of prejudice ($M = 1.9$, $SD = .96$), and had on average a moderate familiarity with gay men ($M = 3.03$, $SD = 1.35$). Also, participants indicated to have heard 1 (65% of participants) or 2 (35% of participants) gay speakers during the experiment indicating that most participants perceived at least one gay-sounding speaker as gay (see Fasoli et al., 2017 for similar results). The average percentage of accuracy for comprehension questions was 92.29%, with similar values across conditions (91.78%, $SE = 4.64$ and 90.82%, $SE = 4.89$ for gay-sounding speakers uttering stereotypically masculine and feminine professions, respectively; 92.51%, $SE = 4.45$ and 90.82%, $SE = 4.89$ for straight-sounding speakers referring to stereotypically masculine and feminine professions, respectively); this indicated that, overall, participants carefully listened to the sentences.[1]

3.2. ERP analyses and results

Grand-averages inspection (Fig. 1) shows that our experimental manipulations seem to affect the ERP waveform from about 200 ms after the target word onset. The deflections between stereotypically masculine and feminine profession conditions are broadly distributed over the scalp but larger on frontal sites. Aside from topographical considerations, the effect is unlikely to be interpreted in terms of a modulation of the N400 component since the polarity of the effect is opposite from what would be expected: the two conditions in which the speaker's voice and message are stereotype-consistent (i.e., stereotypically masculine professions spoken by straight-sounding voices and stereotypically feminine professions spoken by gay-sounding voices) show a negativity with respect to the stereotype-inconsistent conditions. The time development of these effects ranges from 200 ms up to the end of the analyzed epoch, apparently with an earlier onset for straight-sounding (peaking at ~500 ms) than for gay-sounding voices (peaking at around ~550 ms).

To test the statistical significance of this effect, and following van Berkum et al. (2008), a 200–700 ms time window was selected. To reduce the Degrees of Freedom of topographic factors in the statistical analysis, single channel ERP data were averaged in different groups of sites for which the visual inspection showed a relative spatial homogeneity of the effects (the same approach was also used by Van Berkum et al. (2008)). To qualify the distribution of the effect on the scalp and cover all areas of the scalp with sufficient spatial resolution, two repeated (participants) measures ANOVAs were performed, that is one involving groups on the central line, and one involving groups on the lateral sites of the scalp. The analysis of the central line was run on three groups: Frontal (F1, Fz, F2, FC1, FCz,

[1] Note that, in the ERP analysis, no epoch was excluded based on the accuracy of comprehension questions. In fact, the questions only involved the 8% of trial, and an accuracy of 90% means that only less than 1% of epochs would have been rejected on this basis, with a consequent negligeable impact on the ERP data. Moreover, errors cannot be surely linked to a "wrong" or "atypical" on-line processing of the target words, but also/mainly to post-sentence ruminative and task-related processes. Rejecting trials with wrong off-line answers would thus results only in a (rather arbitrary) penalty for trials that would have probably been processed correctly.

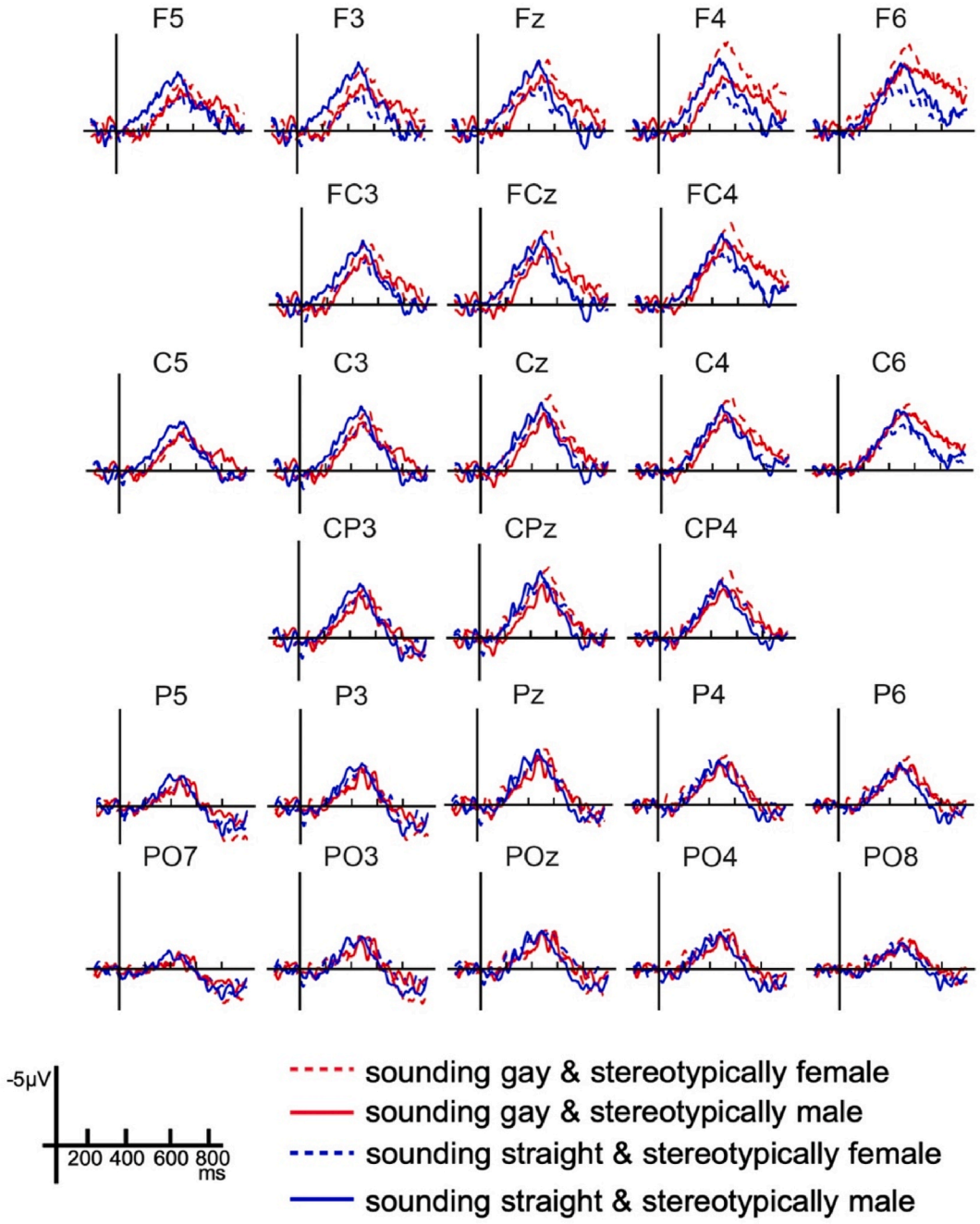


Fig. 1. ERPs waveforms for the different conditions at 26 representative electrodes. ERPs are time-locked to the profession onset.

FC2), Central (C1, Cz, C2, CP1, CPz, CP2), and Parietal (P1, Pz, P2, POz, O1, Oz, O2); in the 2 x 2 x 3 ANOVA we contrasted the factors: Speaker (gay-vs. straight-sounding), profession (stereotypically masculine vs. feminine), and the topographic factor longitude (frontal vs. central vs. parietal). The analysis of lateral sites was run on six groups: Left frontal (F3, F5, F7, FC3, FC5, FT7), Left central (C3, C5, CP3, CP5, TP7), Left parietal (P3, P5, P7, PO3, PO7), Right frontal (F4, F6, F8, FC4, FC6, FT8), Right central (C4, C6, CP4, CP6, TP8), Right parietal (P4, P6, P8, PO4, PO8). In the ANOVA we thus contrasted the following factors: Speaker (gay-vs. straight-sounding), profession (stereotypically masculine vs. feminine), and the topographic factors of longitude (frontal vs. central vs. parietal) and hemisphere (left vs. right). In both analyses, all factors were within participants. Greenhouse-Geisser correction was applied when prescribed (only corrected p values are reported).

The analysis of the central line showed a main effect of longitude (F [2, 70] = 41.33, $p < .001$, $\eta G^2 = .06$); moreover, there was a significant three-way interaction between speaker, profession, and longitude (F [2, 70] = 5.29, $p = .02$, $\eta G^2 = .004$). To better interpret

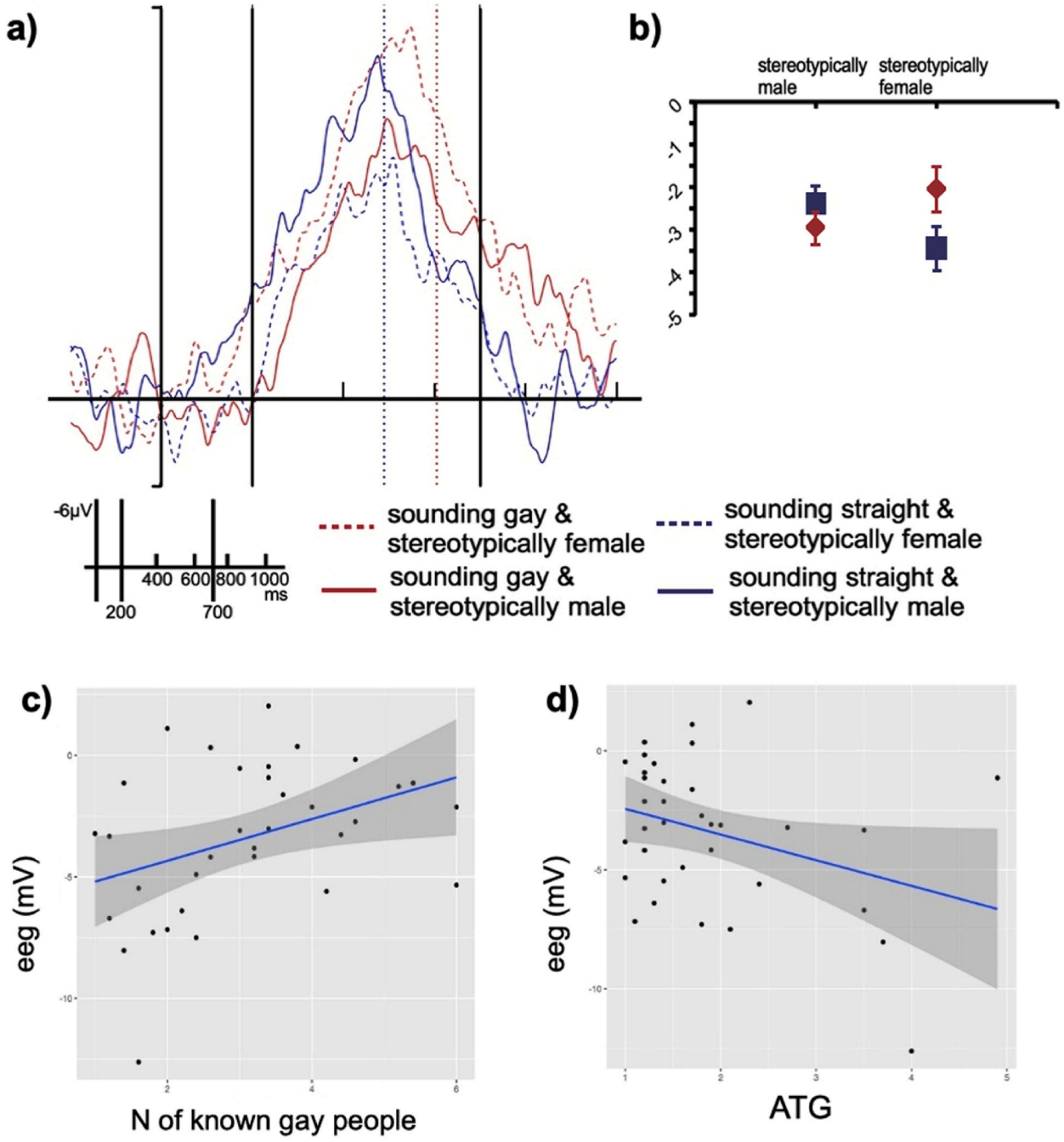


Fig. 2. The significant interaction at the frontal midline group (average of F1, Fz, F2, FC2, Fcz, FC2) and correlations with self-reported measures. a) Grand-average waveforms showing the significant interaction; the first vertical line indicates the profession onset; the following two vertical lines indicate the boundaries of the time windows for the analyses (200–700 ms); the two dashed vertical lines indicate the end of the profession for sounding straight (blue) and sounding gay (red) speakers. Negative voltages are plotted up. b) Mean voltage (with error bars) of the four conditions in the significant time window; blue and red indicates gay-sounding and heterosexual sounding voices, respectively. c) Correlation between the number of gay people known by participants and the difference between the amplitude of the negativity for the stereotypically feminine and masculine professions pronounced by gay-sounding voices. d) Correlation between the score at the prejudice scale and the difference between the amplitude of the negativity for the stereotypically female and male professions pronounced by gay-sounding voices. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

this 3-way interaction, data were split for the topographic factor; the two-way ANOVAs showed that the interaction between speaker and profession was significant frontally ($F[1, 35] = 4.78$, $p = .03$, $\eta G^2 = .03$), but neither centrally ($F = 1.77$, $p > .1$), nor posteriorly ($F < 1$, $p > .6$). Paired *t*-test comparisons (Bonferroni corrected) at frontal sites failed to show any significant difference (both *ps* > .1). Qualitatively, when uttering stereotypically feminine professions, self-identified gay-sounding speakers were associated with a larger negativity (*MEAN* = −3.44 μV, *SE* = .52) than straight-sounding speakers (*MEAN* = - 2.05 μV, *SE* = .53; see Fig. 2b).

The analysis of the lateral sites showed significant effects of the topographic factors (main effects of longitude: $F[2, 70] = 40.71$, $p < .001$, $\eta G^2 = .04$ and hemisphere: $F[1, 35] = 66.93$, $p < .001$, $\eta G^2 = .03$). The two-way interaction between speaker and profession was significant ($F[1, 35] = 6.84$, $p = .01$, $\eta G2 = .03$). The three-way interaction between speaker, profession, and longitude was also significant ($F[2, 70] = 4.53$, $p = .03$, $\eta G^2 = .003$). As done above, the interaction was inspected by splitting the data for the topographic factor. The two-way interaction between speaker and profession failed to reach conventional significance at any site (frontal sites: $F = 2.42$, $p > .1$; central sites: $F = 1.23$, $p > .2$; posterior sites: $F < 1$, $p > .8$). No further effect of the analysis reached significance (all *Fs* < 3, *ps* > .09).

A further analysis was run on the 700–1000ms time window, in order to explore possible speakers-related differences in a later time window. Two ANOVAs, one for central and one for lateral sites, were run with the same factors as the previous ones. The ANOVA on central sites showed a main effect of longitude ($F(2,70) = 19.14$, $p < .001$, $\eta G^2 = .02$) and a two-way interaction between speaker and longitude ($F(2,70) = 6.87$, $p = .006$, $\eta G^2 = .005$). The inspection of the interaction (as above, by splitting the interaction for the topographic factor) showed that the effect of speaker was significant at frontal sites ($F(1,35) = 4.35$, $p = .04$, $\eta G^2 = .03$), with gay sounding speakers associated with a reduced positivity compared to straight sounding speakers, but not at posterior or central sites (both *Fs* < 1, *ps* > .3). The ANOVA on the lateral sites showed significant effects of topographic factors (longitude: $F(2,70) = 38.78$, $p < .001$, $\eta G^2 = .05$; hemisphere: $F(1,35) = 29.72$, $p = .02$). Moreover, speaker interacted with both longitude ($F(2,70) = 6.75$, $p = .01$, $\eta G^2 = .004$) and hemisphere ($F(1,35) = 4.15$, $p = .04$, $\eta G^2 = .003$). The inspection of the interactions (as above) showed that the effect of speaker was only significant at frontal sites ($F(1,35) = 6.08$, $p = .01$, $\eta G^2 = .04$; central sites: $F = 4.06$, $p > .05$, posterior sites: $F < 1$, $p > .6$), with again gay sounding speakers associated with a reduced positivity compared to straight sounding speakers. No further effect was significant (all *Fs* < 3.9, *ps* > .05).[2]

3.3. Correlation analyses

To investigate whether the electrophysiological response to stereotypical stimuli was related to individual differences, correlations were run between the individuals' amplitude of the ERP effect – measured as the difference between the amplitude of the negativity for the stereotypically feminine professions uttered by gay-sounding and straight-sounding speakers, on the Frontal site group within the time interval of the ANOVA (200–700ms) – and sexual prejudice and familiarity with gay men scores. A positive correlation emerged between the size of the frontal ERP effect and the familiarity with gay men ($r = .38$, $p = .02$, Fig. 2b), indicating that the higher familiarity, the smaller the negativity. Also, a negative correlation emerged between the size of the ERP effect and sexual prejudice ($r = -.33$, $p = .04$), indicating that the more positive the attitude toward gay men, the smaller the negativity (Fig. 2c).

4. General discussion

The results of our experiment show that the speaker-meaning inconsistency affects sentence processing rather quickly, as indicated by the clear interaction between the profession-related stereotypes and the voice-based perception of the speakers' sexual orientation. The ERP pattern on the frontal region of the scalp showed a numerically larger negativity for stereotypically feminine professions when uttered by gay-sounding rather than by straight-sounding speakers peaking ~500 ms after the inconsistency presentation. Moreover, correlation analyses indicated that the extent to which the speaker-meaning inconsistency affected sentence processing is related to both the listeners' prejudice and familiarity with gay men: A larger negative effect was observed in participants with higher levels of prejudice and lower familiarity with gay men.

The effect of speaker-meaning inconsistency elicited a negativity that was unexpectedly maximal in anterior scalp regions. The timing of the effect suggests that the information about the speaker is quickly inferred and used by listeners to process and interpret sentences. This is in contrast with the view that a social category inferred from voice is taken into account only at a late stage of processing (Lattner & Friederici, 2003). Although the timing of the effect we reported here is fully compatible with that reported by van Berkum et al. (2008) – and would thus be compatible with a modulation of the N400 component – the anterior topographic distribution of the effect as well as its polarity does not allow an interpretation in terms of a N400 modulation. For straight-sounding speakers, we expected a larger N400 for stereotypically feminine than masculine professions, but a numerical difference was found in the opposite direction; the same reverse pattern (numerically) emerged for gay-sounding speakers. Also, a numerically larger negativity emerged for feminine professions when uttered by gay- than by straight-sounding speakers, corresponding to a larger negativity to the somewhat more congruent condition (within the assumption that feminine professions are stereotypically associated with gay

[2] To rule out any possible effect of participants' gender, we ran the same analyses on the central and lateral sites adding participants' gender as a between-participants factor. The results of the new analyses were in line with those reported in the main text. The only significant effects of participants' gender was in interaction with channel (central sites: $p = .03$; lateral sites: $p = .02$). All the other effects of participants' gender were not significant (central sites: all *ps* >. 1; lateral sites: all *ps* > .08). Hence, participants' gender does not contribute to the interaction between speaker and profession mainly visible at frontal sites.

men). In terms of topography, the N400 observed in sentence comprehension studies is typically maximal around the vertex (Cz), and often visible also on posterior sites. Here, instead, the effect of our manipulation strongly interacts with topographical factors, being clearly detectable more anteriorly than Cz.

Therefore, our results are in contrast even with those reported by van Berkum and colleagues (2008) who found a typical N400 (i.e., with a posterior distribution) for speaker-message stereotypically inconsistencies. There are at least two reasons that may explain this difference. First, our stimuli always referred to speakers' involvement in professions and the content was considered plausible. Instead, van Berkum et al.'s (2008) stimuli consisted of a heterogeneous set of sentences including content that can be perceived as unlikely and/or socially unacceptable (e.g., a baby voice saying "*Every evening I drink some wine before I go to sleep*"). The use of such sentences in van Berkum et al. (2008) may have enhanced the chance of finding a typical N400, as the retrieval of semantic-conceptual knowledge would have been harder because of the 'anomalous' content. A second important difference is related to the type of social information and stereotypes under investigation. Sexual orientation is a more "ambiguous" social category than gender and age (Tskhay & Rule, 2013). Listeners can be hesitant to categorize speakers as gay or straight (Sulpizio et al., 2015; for categorization fluency see Masi & Fasoli, 2022). When a social category is easily recognizable and people are not hesitant in making such categorization, like in the case of gender (e.g., Belin et al., 2004; Linville, 1996; Munson & Babel, 2007), the connections between the category and the related semantic knowledge may be stronger. Indeed, when social categorization occurs, stereotypes are quickly activated (Devine, 1989) and individuals are more confident about their perceptions. It follows that a bigger effort in semantic processing, triggering a typical N400, may be more likely when an unexpected category-content association occurs for salient social categories (e.g., gender, age) rather than ambiguous ones (e.g., sexual orientation).

The unexpected frontal effect we observed needs interpretation. Given such an effect was unexpected and there is no clear link to such a component in the literature of sentence comprehension, its interpretation will be clearly rather speculative, also considering psychophyisiological literature not directly linked to sentence processing.

A modulation of frontal negativity has been repeatedly reported in studies investigating stereotype incongruency (with pictorial and linguistic materials) (e.g., Brusa et al., 2021; Hehman et al., 2014; Pélissier & Ferragne, 2022), which suggest a sensitivity of this component to violation of social representations. Interestingly, in a race categorization task with White participants, Ito and Tomelleri (2017) reported a larger frontal negativity for White- (i.e., in-group congruent) than Black-face categorization, and associated this effect with the processing of implicit stereotyping (for a similar pattern of larger frontal negativity for in-group than out-group faces in a memory recognition task and its interpretation as the access to social information, see also Proverbio et al., 2020).

In social neuroscience, frontal negativities have been often associated with regulatory mechanisms involved in conflict resolution. West and Alain (2000) suggested that this component reflects the suppression of a conceptual processing system during incongruent trials (for a similar interpretation, see also van Hooff et al., 2008). A modulation of this component has also been observed by Bartholow et al. (2006), who found that negativity was larger when participants inhibited stereotype-consistent responses in a racial bias task, and its amplitude correlated with better self-regulation of bias, suggesting a link to regulative control mechanisms.

Our results may fit with the above interpretation of frontal negativity as an index of regulatory mechanisms aimed at controlling for semantic knowledge and suppress stereotypical conceptual representations. It has been suggested that frontal negativity reflects the activity of the prefrontal cortex (see Amodio et al., 2013), a structure that plays a key role in regulative control processes. These control processes may act when automatic semantic retrieval is not enough to ensure the generation of a suited representation of the message meaning (Lambon Ralph et al., 2017). In line with such involvement of semantic control in situations needing the suppression of over-learned responses/association, and under the assumption that social cognition may be considered as a case of semantic cognition (Binney & Ramsey, 2020), our results suggest that the inconsistency between the stereotypical knowledge elicited by the speaker's voice and the knowledge elicited by the message needs the involvement of semantic control to be overcome. Specifically, the larger negativity for the stereotypically feminine professions spoken by gay- than straight-sounding speakers may be the result of the listeners' effort to control and suppress the stereotypical conceptual representation of gay people with stereotypically feminine traits activated by a gay-sounding voice. Gay men are likely to be associated with feminine characteristics (Blashill & Powlishta, 2009; Kite & Deaux, 1987). As a consequence, a sounding-gay speaker quickly and automatically triggers the projection of its stereotypical properties into the speaker's representation (see Fasoli et al., 2017). Thus, control mechanisms would be engaged to control for this quick stereotypical characterization of the speaker.

It is worth noting that the amplitude of the negativity was related to both the listeners' familiarity with gay men and their prejudice – with a smaller effect for listeners being more familiar with gay men. This result indicates that the person's social knowledge has a direct impact on the extent to which stereotypic expectations affect linguistic meaning in shaping the message representation. Reasonably, those who have more (and possibly more intimate) contact with gay men may be aware of the variety of professions held by gay individuals and may be engaging less with stereotypes (see Pettigrew, 1998). Being less prone to make stereotypical associations and inferences, these listeners do not have conceptual representations to suppress. On the contrary, people who have little or no contact with gay men may rely more on stereotypes as these represent their only knowledge about that social category. In such a case, control mechanisms would help the listeners to control for their stereotypical expectations, which could bias subsequent message interpretation and communication. Contact is related to prejudice since those who have no contact with gay men usually report more negative attitudes toward them (Herek & Glunt, 1993). Findings from Italy (Fasoli et al., 2016), where this current research has been conducted, have indeed shown that sexual prejudice is lower in participants who are emotionally closer to gay people. It is hence plausible that individuals who report more prejudice show the opposite pattern of results than those with more frequent contact with gay men.

A second result emerging from our analysis is the late effect of speaker's identity, with gay speakers being associated with a reduced positivity compared to straight speakers. Many late positive effects are found in the field of sentence comprehension, which have been

described as several components (e.g., P600, LPC, Post Negativity Positivity). These components are typically larger for more complex semantic integration processes (e.g., Canal & Bambini, 2023; DeLong & Kutas, 2020), but their functional interpretation as a function of the topographic distribution (frontal, widely distributed or posterior) is still debated. However, the effect we report seems hardly ascribable to any of these components, since they are typically larger for conditions that are more difficult to integrate, as in the case of the P600 reported by Lattner and Friederici (2003), which also had a posterior distribution. This does not seem our case, in which a difference only emerged between straight and gay speakers, without being modulated by stereotypical congruency. Tentatively, our late effect might be related to some more generic aspects of control, not linked to the specific content of the message. Of course, this is a highly speculative interpretation that needs to be validate by future studies.

In terms of neurocognitive models of sentence comprehension, our results are worthwhile to be discussed within the Memory, Unification and Control framework (MUC; Hagoort, 2005, 2013). One salient feature distinguishing Unification from semantic composition is that both linguistic and non-linguistic information (i.e., higher-level pragmatic and communicative information, including context, visual environment, and speaker) are used within the same process whose ERP correlate is the N400 component. On this basis, Hagoort (2017) states that "linguistic and non-linguistic information conspire to determine the interpretation of an utterance on the fly" (p. 2). Our findings cast doubts on the fact that this conspiracy is only played within the Unification component. If our speculative interpretation of the frontal interaction in terms of control processes holds, one may argue that the Control component is also at work in linking the speaker's identity with the information contained in the message. Within the MUC model, the Control component is typically advocated to be involved in turn-taking and target language selection in bilinguals (Hagoort, 2005), but it may be putatively involved in controlling for the active semantic information and selecting or shifting the pragmatic interpretation of the message as a function of the speaker's social identity or communicative context (Hagoort, 2016). To speculate, inconsistency between pragmatic and contextual information and a target word might be on-line processed by different routes. One route implies the modulation of the classic central-posterior N400 (e.g., Otten et al., 2007; Van Berkum et al., 2008), which is in line with the Hagoort's broad concept of semantic composition as Unification. The other route, as suggested by our finding, implies the recruitment of different neural sources that elicit a frontal negativity (Foucart et al., 2015; Otten et al., 2007), possibly linked to the recruitment of additional control processes during comprehension.

4.1. Limitations and future directions

This study is not without limitations. First, the study was not pre-registered and our sample was small and not representative of the population. Indeed, the same involved mostly individuals who identified as cisgender heterosexual women – who usually report more positive attitudes toward and more contact with gay men than cisgender men do (Herek & Glunt, 1993). Future research should consider include participants who vary in terms of gender identities and sexual orientations to examine if similar patterns emerge. Moreover, having a larger sample would be particularly important to capture more variation in levels of prejudice and familiarity and would allow researcher to have more solid evidence concerning correlations that were here assessed on a small number of data points (and for this reason could be considered not much informative).

Second, we used recordings of only two speakers for each sexual orientation and since variations among speakers occur (Levon, 2007; Kachel, Simpson, & Steffens, 2018), this limits the generalizability of our results to other speakers. Also, our speakers identified as men but varied in their perceived gender typicality in relation to their sexual orientation. They were clearly distinguished in terms of 'gender prototypicality' (i.e., perceived masculinity), but less clearly differentiated in terms of sexuality. This potentially means that the effects might emerge among speakers who sound more or less masculine regardless of the perceived sexual orientation. It is difficult to disentangle the two aspects as they go hand in hand (see Munson et al., 2006; Painter et al., 2024) and because listeners are often hesitant to categorize others as gay (Sulpizio et al., 2015) as this is seen as stigmatizing (see Alt et al., 2020). Future studies should engage speakers more clearly perceived as gay or manipulate information about the speakers. For instance, using speakers who sound more or less masculine and are explicitly introduced as either gay or straight could help assess message content processing when expectations are violated (see Gowen & Britt, 2006). Moreover, we referred to a binary concept of gender, but gender identity varies and may not overlap with the sex assigned at birth. Hence, it would be important to consider how gender fluidity and gender identities expressed by voice affect gender stereotypical messages (see Zimman, 2018). For instance, vocal femininity has been found to affect the perception of competence (Ko et al., 2009) and, thus, this vocal feature could affect the interpretation of competence-based messages.

Third, we did not assess or manipulate the plausibility and social acceptability of our stimuli. Differences in methodological and stimuli choices may affect how the inconsistency between speaker and message impacts ERPs, modulating the N400 in van Berkum et al.'s (2008) study and frontal negativity in the present one, and the P600 in Lattner and Friederici (2003) and Foucart et al.'s (2015) study. Hence, future studies should investigate what methodological choices are responsible of the differences across studies.

5. Conclusion

This experiment investigated the time course of speaker-meaning integration during sentence comprehension. We focused on when listeners integrate information about the social identity extracted by the speaker voice with the message uttered by the speaker. Our results suggest that this integration occurs very quickly and that information about the speaker's social identity is used to build the message meaning. Moreover, during this process, control processes can be engaged online to deal with inconsistencies, but their level of involvement seems to be associated with listeners' individual differences. Taken together, our results contribute to shed light on the complexity of communication that involves stereotype-based messages, the social categorization of speakers, and listeners'

characteristics.

CRediT authorship contribution statement

Simone Sulpizio: Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fabio Fasoli:** Writing – original draft, Supervision, Methodology, Conceptualization. **Gaia Lapomarda:** Writing – review & editing, Investigation, Data curation. **Francesco Vespignani:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was partially supported by a grant of the Fondazione Cassa di Risparmio di Trento e Rovereto.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jneuroling.2025.101248.

Data availability

Data will be made available on request.

References

Alt, N. P., Lick, D. J., & Johnson, K. L. (2020). The straight categorization bias: A motivated and altruistic reasoning account. *Journal of Personality and Social Psychology, 119*, 1266.

Amodio, D. M., Bartholow, B. D., & Ito, T. A. (2013). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience, 9*, 385–393.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology, 90*, 272–287.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129–135.

Binney, R. J., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience & Biobehavioral Reviews, 112*, 28–38.

Blashill, A. J., & Powlishta, K. K. (2009). Gay stereotypes: The use of sexual orientation as a cue for gender-related attributes. *Sex Roles, 61*, 783–793.

Brambilla, M., Riva, P., & Rule, N. O. (2013). Familiarity increases the accuracy of categorizing male sexual orientation. *Personality and Individual Differences, 55*, 193–195.

Brusa, A., Bordone, G., & Proverbio, A. M. (2021). Measuring implicit mental representations related to ethnic stereotypes with ERPs: An exploratory study. *Neuropsychologia, 155*, Article 107808.

Canal, P., & Bambini, V. (2023). Pragmatics electrifiyed. In M. Grimaldi, E. Brattico, & Y. Shtyrov (Eds.), *Language electrified: Principles, methods, and future perspectives of investigation* (Vol. 202, pp. 583–612). Springer Nature, 2023.

Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology, 6*, 1953.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience, 7*, 764–770.

Crepaldi, D., Keuleers, E., Mandera, P., & Brysbaert, M. (2013). SUBTLEX-IT. Retrieved from http://crr.ugent.be/subtlex-it/.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences, 27*(4), 801–804.

DeLong, K. A., & Kutas, M. (2020). Comprehending surprising sentences: Sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Language, Cognition and Neuroscience, 35*, 1044–1063.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5.

Diamond, L. M. (2016). Sexual fluidity in male and females. *Current Sexual Health Reports, 8*, 249–256.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69*, 275–298.

Fasoli, F., & Formanowicz, M. (2024). Can agentic messages help? Linguistic strategies to counteract voice-based sexual orientation discrimination. *British Journal of Social Psychology*. https://doi.org/10.1111/bjso.12739.

Fasoli, F., Hegarty, P., & Frost, D. M. (2021). Stigmatization of 'gay-sounding'voices: The role of heterosexual, lesbian, and gay individuals' essentialist beliefs. *British Journal of Social Psychology, 60*, 826–850.

Fasoli, F., & Maass, A. (2018). Voice and prejudice: The social costs of auditory gaydar. *Atlantic Journal of Communication, 26*, 98–110.

Fasoli, F., Maass, A., & Berghella, L. (2023). Who has a better auditory gaydar? Sexual orientation categorization by heterosexual and lesbian, gay and bisexual people. *Journal of Homosexuality, 70*, 876–899. interpretation. Journal of Language and Social Psychology, 39, 653-677.

Fasoli, F., Maass, A., Karniol, R., & Antonio, R. (2020). Voice changes meaning: The role of gay-versus straight-sounding voices in sentence interpretation. *Journal of Language and Social Psychology, 39*, 653–677.

Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay-and lesbian-sounding auditory cues elicit stereotyping and discrimination. *Archives of Sexual Behavior, 46*, 1261–1277.

Fasoli, F., Maass, A., & Sulpizio, S. (2018). Stereotypical disease inferences from gay/lesbian vs. heterosexual voice. *Journal of Homosexuality, 65*, 990–1014.

Fasoli, F., Paladino, M. P., & Sulpizio, S. (2016). Conosco persone omosessuali: Una ricerca esplorativa degli effetti del contatto sull'omofobia e sostegno ai diritti LGBT in Italia. *Psicologia Sociale, 11*, 263–286.
Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia, 75*, 291–303.
Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech, 69*, 30–57.
Gowen, C. W., & Britt, T. W. (2006). The interactive effects of homosexual speech and sexual orientation on the stigmatization of men: Evidence for expectancy violation theory. *Journal of Language and Social Psychology, 25*, 437–456.
Grey, J. A., Robinson, B. B. E., Coleman, E., & Bockting, W. O. (2013). A systematic review of instruments that measure attitudes toward homosexual men. *The Journal of Sex Research, 50*, 329–352.
Hagoort, P. (2005). On broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*, 416–423.
Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language* (pp. 339–347). Academic Press.
Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience & Biobehavioral Reviews, 81*, 194–214.
Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In *The cognitive neurosciences* (4th ed., pp. 819–836). MIT press.
Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–441.
Hehman, E., Volpert, H. I., & Simons, R. F. (2014). The N400 as an index of racial stereotype accessibility. *Social Cognitive and Affective Neuroscience, 9*(4), 544–552.
Herek, G. M. (1996). Why tell if you're not asked? Self-Disclosure, intergroup contact, and heterosexuals' attitudes toward lesbian and gay men. In G. M. Herek, J. Jobe, & R. Carney (Eds.), *Out of force: Sexual orientation and the military* (pp. 197–225). Chicago: University of Chicago Press.
Herek, G. M., & Glunt, E. K. (1993). Interpersonal contact and heterosexuals' attitudes toward gay men: Results from a national survey. *The Journal of Sex Research, 30*, 239–244.
Herek, G. M., & McLemore, K. A. (1998). Attitudes toward lesbians and gay men scale. In T. D. Fisher, C. M. Davis, & W. L. Yarber (Eds.), *Handbook of sexuality-related measures* (pp. 392–394). Thousand Oaks, CA: Sage Publications.
Huic, A., Jelic, M., & Kamenov, Z. (2018). Essentialist beliefs about homosexuality predict positive and negative behavioral intentions toward lesbian women and gay men. *Journal of Homosexuality, 65*, 1631–1655.
Ito, T. A., & Tomelleri, S. (2017). Seeing is not stereotyping: The functional independence of categorization and stereotype activation. *Social Cognitive and Affective Neuroscience, 12*, 758–764.
Jacobs, A. M. (2015). Towards a neurocognitive poetics model of literary reading. *Cognitive neuroscience of natural language use*, 135–159.
Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: In group preferences and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin, 30*, 629–642.
Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology, 73*, 55–79.
Kachel, S., Simpson, A. P., & Steffens, M. C. (2018). "Do I sound straight?": Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men's speech. *Journal of Speech, Language, and Hearing Research, 61*, 1560–1578.
Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly, 11*, 83–96.
Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin, 35*, 198–211.
Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*, 12–35.
Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*, 42–55.
Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience Letters, 339*, 191–194.
Lau, E., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra: Psychology, 2*, 13.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933.
Lehavot, K., & Lambert, A. J. (2007). Toward a greater understanding of antigay prejudice: On the role of sexual orientation and gender role violation. *Basic and Applied Social Psychology, 29*(3), 279–292.
Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society, 36*, 533–554.
Lick, D. J., & Johnson, K. L. (2016). Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments. *Journal of Pesonality and Social Psychology, 110*, 801.
Linville, S. E. (1996). The sound of senescence. *Journal of Voice, 10*, 190–200.
Lyons, A., Clark, A. E., Kashima, Y., & Kurz, T. (2007). Cultural dynamics of stereotypes: Social network processes and the perpetuation of stereotypes. In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the forma- tion, maintenance, and transformation of stereotypes*. New York: Psychology Press.
Lyons, A., & Kashima, Y. (2003). How are stereotypes maintained through communication? The influence of stereotype sharedness. *Journal of Personality and Social Psychology, 85*(6), 989–1005. https://doi.org/10.1037/0022-3514.85.6.989
Mack, S., & Munson, B. (2012). The influence of/s/quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp'stereotype. *Journal of Phonetics, 40*, 198–212.
Masi, M., & Fasoli, F. (2022). When fluency matters: The interplay between categorization fluency and gender atypicality on gaydar judgments. *Journal of Language and Social Psychology, 41*, 746–759.
Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., … Valdrova, J. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods, 46*, 841–871.
Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech, 50*, 125–142.
Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass, 1*, 416–449.
Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics, 34*, 202–240.
Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition, 25*, 273–285.
Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8*, 89.
Painter, D., Fasoli, F., & Sulpizio, S. (2024). The impact of stimuli length and analytic method on auditory "gaydar" research. *Journal of Voice, 38*, 246–e1.
Pélissier, M., & Ferragne, E. (2022). The N400 reveals implicit accent-induced prejudice. *Speech Communication, 137*, 114–126.
Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.
Preuß, S., Ottenstein, C., Kachel, S., & Steffens, M. C. (2020). Using scenarios for measuring the affective and behavioral components of attitudes toward lesbians and gay men: Validation of the SABA scale. *Archives of Sexual Behavior, 49*, 1645–1669.
Proverbio, A. M., Parietti, N., & De Benedetto, F. (2020). No other race effect (ORE) for infant face recognition: A memory task. *Neuropsychologia, 141*, Article 107439.
Salvati, M., Piumatti, G., Giacomantonio, M., & Baiocco, R. (2019). Gender stereotypes and contact with gay men and lesbians: The mediational role of sexism and homonegativity. *Journal of Community & Applied Social Psychology, 29*, 461–473.
Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS One, 7*(12).

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society, 32*, 329–350.
Sulpizio, S., Fasoli, F., Antonio, R., Eyssel, F., Paladino, M. P., & Diehl, C. (2020). Auditory gaydar: Perception of sexual orientation based on female voice. *Language and Speech, 63*, 184–206.
Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PLoS One, 10*, Article e0128882.
Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups a review and meta-analysis. *Personality and Social Psychology Review, 17*, 72–86.
Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.
Van Hooff, J. C., Dietz, K. C., Sharma, D., & Bowman, H. (2008). Neural correlates of intrusion of emotion words in a modified Stroop task. *International Journal of Psychophysiology, 67*, 23–34.
West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology, 37*, 179–189.
Zimman, L. (2013). Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men. *Journal of Language and Sexuality, 2*, 1–39.
Zimman, L. (2018). Transgender voices: Insights on identity, embodiment, and the gender of the voice. *Language and Linguistics Compass, 12*, Article e12284.