The Impact of Stimuli Length and Analytic Method on Auditory 'Gaydar' Research

Daniel Painter[1], Fabio Fasoli[1] & Simone Sulpizio[2]

[1] University of Surrey

[2] University of Milano-Bicocca

Corresponding author:

Dr Fabio Fasoli

School of Psychology, University of Surrey,

Stag Hill Campus, GU2 7XH, Guildford, United Kingdom

Email: f.fasoli@surrey.ac.uk

Acknowledgment. This research project was supported by the Doctoral College at the University of Surrey who awarded a summer internship to the first author.

Abstract

Keywords: sexual orientation, auditory gaydar, voice, measurements

1. Introduction

Listening to someone's voice allows us to quickly categorize them as member of a given social category (e.g., gender) and also to form a first impression[1, 2, 3]. However, the amount of information we obtain from vocal cues varies across situations. We can listen to someone speaking for a few seconds, for instance when hearing them saying 'hello', or for longer, like when having a conversation with someone on the phone. Does this affect our perception and categorization of the speaker?

We address this question here by focusing on sexual orientation (henceforth SO). Recent research has shown that individuals categorize others' SO quickly after hearing them speaking [4, 5]. SO is defined as an 'ambiguous' category [6]. Indeed, compared to other 'obvious' categories that are marked by perceptual and vocal features (e.g., age), SO cannot be assured until the speaker self-discloses. Moreover, SO is often conceptualized as binary fixed categories (e.g., gay/heterosexual), but it actually lies on a continuum [7, 8]. All these aspects make SO an interesting category to consider when voice is concerned. In such a specific case, the amount of vocal information that is available, namely whether we are exposed to a short (e.g., just a word) or long vocal information (e.g., one or more sentences), could play a role on the SO judgements that listeners make. This research will examine the issue of stimuli length while considering differences that may occur when SO judgments are assessed on binary categories (gay/lesbian vs. heterosexual) or Kinsey-like (heterosexual-gay/lesbian rating) scales.

1.1. Auditory Gaydar

Research has shown that voice is considered a SO cue[9] and that listeners use vocal cues to judge a speaker's SO, a phenomenon called 'auditory gaydar' [4, 5]. Hence, this process consists of listeners' voice-based categorization of a speaker's SO. It has been suggested that accuracy

for auditory gaydar is usually above the chance level, around 63%, but far from being perfect[5]. Indeed, the seminal work by Gaudio (1994) and results by Linville (1998) indicated that heterosexual listeners were overall accurate in recognizing male speakers' who self-identified as either gay or heterosexual. Such findings also found support in a more recent study by Tracy, Bainter, and Satariano (2015). However, other studies[10, 11] have provided contrasting evidence on gaydar accuracy. For instance, Sulpizio et al., (2015) showed that auditory gaydar judgments are often inaccurate and that listeners cluster male speakers into gay- and heterosexual-sounding groups that do not always correspond to the way the speakers self-identify. These contrasting findings can also be observed among the limited studies conducted on female speakers. While auditory gaydar accuracy for female speakers emerged in some studies[12], inaccuracy was found in other research[11, 13, 14]. Variation in the way both lesbian and heterosexual speakers are perceived[15] suggests that only a minority of female speakers are perceived as lesbian-sounding and this occurs regardless of their actual SO.

The inaccuracy observed in some auditory gaydar studies could be explained by the so-called 'straight categorization bias'[16], namely the tendency to categorize all the targets as heterosexual unless there are gendered cues suggesting that the targets are less prototypical of their gender. Indeed, it has been shown that it is not knowing that 50% of the voice sample or that 10% of the general population is gay, but rather the presence of gendered cues that influences heterosexuals' SO judgments[16]. This effect is in line with the gender inversion theory[17] suggesting that SO perception is associated with perceived gender atypicality and with the fact that SO inaccuracy is often linked to the absence of gendered cues[18]. When voice is concerned, speakers who are perceived to be lesbian/gay are also perceived as sounding less

gender typical[12, 19, 20]. Such stereotype is even present when speakers' self-evaluations are involved[21, 22].

Studies have also focused on acoustic cues in lesbian and gay (henceforth LG) and heterosexual voices, and on which acoustic cues are related to perceived SO and speech stereotypes[23]. Overall, there is agreement that certain cues (like /s/ and mean fundamental frequency) occurring in speakers' voices reflect specific voice stereotypes that are, in turn, used to categorize speakers as LG[24, 25]. For instance, male speakers whose voice involves lisping are more likely to be categorized as gay[25], while a more monotone speech in female speaker is associated with perception of the speaker as lesbian[26]. Recently, it has been showed that, to a certain degree, there is an overlap between acoustic cues signaling actual and stereotype-based SO, and that exposure to SO stereotypical voices make auditory gaydar salient[23].

1.2. Auditory Gaydar and Methodological Issues

Auditory gaydar research has provided mixed results on whether speakers are accurate or not when judging SO. Thus, it became particularly important to examine what factors could contribute to increase or decrease accuracy.

A first factor refers to the type of stimuli and length of the exposure. First impression leads to quick and automatic social categorization that can however be revised depending on the type of information available[27, 28]. The question of whether type of stimuli and stimulus length can affect impression formation has been mostly addressed in the context of visual information[29]. However, recently, scholars have investigated whether voice-based personality judgments change depending on the stimulus length (a single word or a full sentence) or whether the information listeners are exposed to is socially-ambiguous (meaningless word) or socially-relevant (meaningful content).[30] Although results has shown a general consistency between

listeners' personality judgments regardless of the type of audio information, there were instances in which stimulus length mattered[31]. In the context of visual gaydar, studies have shown that both length of exposure and type of visual stimuli can affect SO judgments. In particular, Rule and Ambady (2008) found that SO judgments were inaccurate when visual stimuli were presented for 33 milliseconds, but that 50 milliseconds or longer exposure was enough for participants to accurately detect SO from male faces. Also, comparing participant's responses to full-face, internal features, and eyes-only stimuli found that as the amount of information decreases the accuracy of an individual's gaydar will increase[33]. This raises the question of whether the length of audio stimulus and linguistic information would also affect SO judgments and accuracy when voice is concerned. Research on auditory gaydar implemented different types of stimuli that varied both in terms of length and content (e.g., single phoneme, words, sentences, or texts; for an overview, see Table 1). For instance, initial studies [10, 33, 34] involved long texts that varied in content (e.g., dramatic text), while more recent studies involved only few words or short neutral sentences[13, 14, 19, 20].

- Table 1 -

The importance of comparing different types of stimuli and stimuli length for auditory gaydar research has been recently put forward[22]. So far, only one study has looked at stimuli length in the context of auditory gaydar[35]. It examined SO judgments when a consonant, a consonant and vowel, or the entire word (4-5 phonemes) was presented. Results showed an overall gaydar accuracy when participants were exposed to single vowels. The accuracy increased when listeners could rely on multiple phonemes such as vowels and stereotypical acoustic cues (e.g., /s/) or a full word. However, notwithstanding the importance of this research, it merely focused on single words or very brief and meaningless stimuli (e.g., single phonemes).

Also, it only considered male speakers leaving unexplored whether stimuli length may affect SO judgments for female speakers too. Hence, we are yet to understand whether stimuli (e.g., words and sentences) that vary both in terms of length and meaningful content affects accuracy and whether speaker gender differences may occur. Longer audio stimuli provide richer and comprehensive information such as multiple vowels and consonants[18, 35] and prosody [36, 37] that influence speakers' SO perception. Following the results by Tracy et al. (2016), one could be put forward that longer audio stimuli may facilitate more considered SO-judgments and therefore accuracy. However, since listening to short stimuli would only allow listeners to form a quick impression[2], and the continuous impression formation literature suggests that automatic social categorization can be revised when additional information becomes available (as it happens in the visual gaydar context[26]), longer stimuli may introduce doubts leading to more 'blurred' SO judgments.

Another aspect that requires attention is how categorization is assessed. Categories are usually preferred as they make social judgments easier[38] but a continuous conceptualization may be helpful too. Introducing individuals with a categorical rather than a continuous description affects the way others perceive and make inferences about them[40]. Such differences may be explained by the fact that presenting categories lead perceivers to overestimate similarity between members of the same group and enhance perceived difference with other group members[41, 42]. A continuous measure allows to detect how similar and prototypical an individual is of a given category while allowing to make a judgment that goes beyond restrictive categories. For instance, listeners' perception of speakers' gender lies on a continuum rather than on distinct categories [43] and hence a continuous variable would be better suited to assess such perception (see gender typicality). Similarly, SO can be conceptualized as a binary (LG vs. heterosexual) or

as a more fluid concept[8]. As a matter of fact, it has demonstrated that describing SO in more fluid terms influence individuals' SO conceptualization moving away from discrete categories[44].

The issue of how SO is operationalized goes hand in hand with how gaydar accuracy is defined (for a discussion see [39]). As shown in Table 1, auditory gaydar studies have used different measures to examine gaydar. Some involved SO binary choices and above chance percentages of correct responses have been considered indexes of gaydar accuracy. Such type of SO categorization allows scholars to assess the *absolute* accuracy, namely the ability to correctly detect the speakers' SO. Other studies used Kinsey-like rating scales that allowed participants to modulate their answers from 'exclusively heterosexual' to 'exclusively gay/lesbian.' In this case, mean differences between LG and heterosexual speakers on Kinsey-like scales were interpreted as gaydar accuracy, even if all means fall on the heterosexual pole of the scale[28]. Such interpretation has been criticized[19]. Indeed, means below the Kinsey-like scale midpoint can illustrate the so-called 'straight categorization bias'[16] indicating that – regardless of actual mean difference - listeners tend to perceive speakers as heterosexual showing an overall inaccuracy and hesitance in categorizing someone as LG[14, 20]. However, this type of measure is important because it allows scholars to assess SO judgments differently than binary choices. Indeed, Kinsey-like scales focus on the *relative* accuracy, namely the ability to differentiate between the two groups of speakers, even when ratings are on the same pole of the scale. Hence, although the two measures have been both used to assess accuracy, they provide two different types of information.

The type of SO measure that is implemented may be particularly important in relation to the type of vocal stimuli in use. Social categories are used to simplify the processing of social information[45] and SO binary choices allow quick decisions on whether the speaker is LG or

heterosexual. This type of measure may facilitate SO judgments when short stimuli, and limited vocal information, are presented. Kinsey-like scales may require individuals to think more and elaborate on SO as a fluid concept[46]. To our knowledge, no auditory gaydar research has measured SO using both binary choices and Kinsey-like scales at the same time.

1.3. Overview

Across three studies we examined whether stimuli length affected SO judgments. We conceptualized length in terms of both duration, amount, and type of lexical information provided. First, we tested heterosexual listeners' accuracy of SO judgments concerning male speakers (Study 1) and then extended it to female speakers (Study 2) when exposed to either a single lexical unit, namely a word, or multiple lexical units forming meaningful sentences. In Study 3, we examined SO judgments when stimuli type was manipulated as the sequential words contained in a sentence: an article, an article and a word, a full sentence. This mimicked the sequential exposure to lexical content listeners are used to and, similarly to previous studies[34], it allowed us to have a short stimulus consisting of few phonemes (i.e., two consonants and a vowel). Moreover, it also allowed us to manipulate the stimulus length while assuring the stimuli came from the same linguistic materials. In all the studies we relied on previously-used measures and examined SO judgments made on a binary choice and on a Kinsey-like scale, respectively. This allowed us to compare our results with previous work, to both explore whether similar patterns of SO judgments emerged across measures and whether both types of SO judgments varied as a function of stimuli type. Moreover, we measured listeners' confidence in their SO-judgments and speakers' perceived gender typicality. Literature has found confidence[47] and masculinity/femininity to be related to SO-judgments[18, 21]. Hence, we tested whether longer

stimuli would trigger an increase in auditory gaydar accuracy, more confidence in SO judgments, and as a stronger perception of LG speakers as gender atypical than shorter stimuli.

2. Study 1

Study 1 aimed to examine the impact of stimuli length when male speakers were concerned. In so doing, Study 1 extended previous work on stimuli duration conducted in the context of visual[48] and auditory gaydar[34] for male targets.

2.1. Method

2.1.1. Participants

89 participants were recruited online via the research platform Prolific Academic and rewarded £0.60. After excluding those participants who did not identify as heterosexual (n = 9), the final sample consisted of 80 (41 women, 38 men, and one who self-defined as 'other'; M_{age} = 37.17, *SD* = 13.80) heterosexual English native speakers. They were all British, predominantly White (91.3%, n = 73), most had a University or college qualification (67.7%, n = 54) and did not indicate any religion (53.7%, n = 43). Participants self-categorized their political orientation as either left-wing (n = 36) or right-wing (n = 44).[a] Participants also reported to know, on average, few gay men (*M* = 3.18, *SD* = 1.85) and reported to be quite unfamiliar with the gay community (*M* = 2.18, *SD* = 0.99; *t*-test against the scale midpoint: $t(79) = 17.24$, $p < .001$).

A GPower[49] sensitivity analysis with power (1 - β) set at 0.80 and α = 05, two measurements and 3 groups indicated that the sample size allowed to detect a small to medium effect size of .17.

[a] Since research on visual gaydar indicated political orientation can play a role in SO judgments with more liberal participants taking longer and reflecting more when categorizing others' SO[54], we measured political orientation on a scale from 1 (*left wing*) to 7 (*right wing*), a direct choice (*left*, *centre*, *right*), and on an 8-item political ideology scale (α = .69; 1 = strongly disagree – 7 = strongly agree)[69]. Participant's average self-reported political orientation (*M* = 3.09, *SD* = 1.86), and political ideology (*M* = 5.04, *SD* = 2.00) were overall more liberal.

2.1.2. Materials

2.1.2.1. Audio stimuli. The stimuli were audio recordings of 5 self-identified gay and 5 self-identified heterosexual British male speakers with English as the first language (age ranging from 20 to 36, without age differences across the two groups of speakers: $M_{heterosexual}$ = 27.40, *SD* = 9.78 vs. M_{gay} = 28.40, *SD* = 9.10, $t(8) = -.17$, $p = .87$). Speakers were recruited among researchers' contacts and University students who received an Amazon vouchers (£10) in exchange for their participation. They were informed that we were recruiting people to record audio stimuli for studies on voice-based impression. Individuals who accepted to participate were recorded in a quiet room using PRAAT installed on a notebook and a Roland R-05HR portable recorder. They were presented with a piece of paper including a list of short neutral sentences among other materials. They first read the sentences out loud to familiarize with them and then recorded them. If a mistake (e.g., mispronouncing) occurred, they were asked to record the sentence again. Next, they completed a short survey where they reported their age, gender, nationality, and SO. Self-reported sexual orientation was assessed both on a categorical variable (gay/lesbian, heterosexual, bisexual, other) or a Kinsey-like scale from 1 (*exclusively heterosexual*) to 7 (*exclusively gay*). We selected speakers who identified as gay or heterosexual. At the end of the recording session, they were fully debriefed and informed about the purpose of the recordings (i.e., studies on auditory gaydar) and, hence, asked to provide final consent for their audio recording use. Audio recordings used in the study referred to speakers who provided such consent.

From the material available, we selected two neutral sentences used in previous auditory gaydar studies[14, 20]. Hence, we created three types of stimuli. In the short condition, participants heard only one word ("Starts", less than a second). We selected such word because it contained

the consonant /s/ that has been found to be associated with perceived SO when it has longer durations, high peak frequency, and skewness[24, 50, 51, 52]. We also selected this word aware that voice-base judgments are similar when a socially-relevant or socially-irrelevant word is presented to listeners[30]. In the medium condition, they heard one sentence ("The dog ran in the park.", 2-3 seconds) and in the long condition, they heard two sentences ("The dog ran in the park. The English course starts on Monday.", 4-5 seconds). These two conditions involved a range of consonants and vowels (for differences in vowels and SO judgments see [11, 19, 23, 50]) as well as allowed for a better understanding of the speaker's prosody.

2.1.2.2. Gender typicality. Participants answered the question "How much do you think this person sounds feminine/masculine?" on a Likert scale from 1 (*completely feminine*) to 7 (*completely masculine*).

2.1.2.3. SO Kinsey-like scale. Participants rated the speaker's SO on a Kinsey-like scale by following this instruction: 'Please, rate the speaker's sexual orientation on the following scale from 1 (*exclusively heterosexual*) to 7 (*exclusively gay*)'.

2.1.2.4. SO binary choice. Participants were instructed to 'indicate if you believe the speaker is heterosexual or gay by choosing one of the two options', the binary choice being: 'heterosexual' vs. 'gay'.

2.1.2.5. Confidence. Participants rated how confident they were about their SO judgments on a Likert scale from 1 (*not at all confident*) to 7 (*absolutely confident*).

2.1.2.6. Gay contacts and familiarity. Participants reported how many gay men they knew on a single item on scale from 1 (*none*) to 7 (*many*) and completed 4 items of the familiarity scale (e.g., 'How often do you interact with openly gay men?'; $\alpha = .89$; Brambilla et al., 2011). Answers were provided on a scale from 1 (*never*) to 5 (*always*).

2.1.3. Procedure

Upon first entering the study, participants were asked to participate in a study on 'perception of sexual orientation', they were presented with the study information and indicated their consent. They were reminded to have their audio devices activated and at an adequate volume. Participants were randomly assigned to one of the three stimulus length conditions: short (n = 28), medium (n = 25), and long (n = 27). Hence, they listened to one audio recording at the time for a total of 10 audio recordings presented in a randomized order. After listening to each audio recording participants completed measures of gender typicality, SO Kinsey-like scale, SO binary choice, and confidence in their SO judgments. Participants were not informed about base rate information since this does not affect SO judgments[16, 20]. Next, they completed the aforementioned scales of political ideology, gay contacts, and gay familiarity. Finally, before being debriefed and providing final consent for data use, participants answered demographic questions (on gender, age, ethnicity, level of education, SO, and first language) and had the opportunity to report any issues with the audio. On average the study lasted 7 minutes.

2.2. Results

We analyzed the data as in previous auditory gaydar research to allow a descriptive comparison. A 2 (Speaker SO: heterosexual vs. gay) x 3 (Stimulus Length: short vs. medium vs. long) repeated measures ANOVA was conducted on each dependent variable assessed with a Likert scale. The first variable was within-participants and the second was between-participants. Pairwise comparisons (Bonferroni correction) were performed in case of significant interactions.

2.2.1. Gender typicality

Ratings were averaged for the gay and heterosexual male speakers. The higher the rating, the more gender typicality attributed to the speakers. A significant main effect of speaker SO was

found, $F(1, 77) = 16.04$, $p < .001$, $\eta_p^2 = .17$, indicating that participants rated gay speakers ($M = 4.98$, $SD = .89$) as less gender typical than heterosexual speakers ($M = 5.35$, $SD = .87$). No other significant effects or interactions were found (Fs < .29, ps > .75).

2.2.2. Sexual Orientation – Kinsey-like Scale

The ratings for gay and heterosexual speakers were averaged, with a higher rating indicating the speaker was perceived as more gay. A significant main effect of Speaker SO, $F(1, 77) = 37.54$, $p < .001$, $\eta_p^2 = .33$, showed that gay speakers ($M = 3.25$, $SD = .79$) were rated as more gay than heterosexual speakers ($M = 2.77$, $SD = 0.77$). Due to the current debate on whether mean difference can be interpreted as accuracy, we also conducted t-test against the scale midpoint (4). Both means were below the scale midpoint (heterosexual speakers: $t(79) = 36.79$, $p < .001$; gay speakers: $t(79) = 32.16$, $p < .001$) suggesting that all speakers tended to be rated on the heterosexual pole of the scale but gay speakers to a less extent than heterosexual speakers.

The main effect of Stimulus length was not significant, but the interaction between Speaker SO and Stimulus Length was, $F(2, 77) = 3.40$, $p = .039$, $\eta_p^2 = .081$. The difference in ratings for gay and heterosexual speakers was significant both in the Short ($M_{gay} = 3.28$, $SD = .91$ vs. $M_{heterosexual} = 2.69$, $SD = .73$; $p < .001$) and Medium condition ($M_{gay} = 3.34$, $SD = .82$ vs. $M_{heterosexual} = 2.69$, $SD = .94$; $p < .001$), but not in the Long condition ($M_{gay} = 3.14$, $SD = .63$ vs. $M_{heterosexual} = 2.78$, $SD = .77$; $p = .155$).

2.2.3. Sexual orientation – Binary choice

Overall, correct answers exceed for heterosexual ($M = 85.50$, $SD = 16.83$) than for gay speakers ($M = 29.50$, $SD = 21.34$). t-tests against the chance level (50%) indicated that accuracy

for heterosexual speakers was above the chance level, $t(79) = 18.86$, $p < .001$, whereas accuracy for gay speakers was below chance, $t(79) = -8.59$, $p < .001$.

To assess accuracy, in this case, we performed a signal detection analysis[53] considering *hit* rates, namely the number of correct categorizations for gay speakers, and *false alarm* rates, namely the number of categorizations of straight speakers as gay, for each participant. Hit rates and false alarms were divided for the number of gay and straight speakers, respectively. Hence, we calculated the *d'* that indicates accuracy or correct discriminability, *d'* of zero indicates guessing while positive scores indicated correct categorization of gay targets. Overall, *d'* indicated poor accuracy ($M = .56$, $SD = 2.01$), *t*-test against zero: $t(79) = 2.50$, $p = .01$. We also calculated the *c* index that represents a measure of response bias. Positive values for *c* indicate a bias in categorizing individuals as heterosexual. Overall, the *c* index showed a response bias toward the heterosexual category, ($M = .81$, $SD = .88$), $t(79) = 8.12$, $p < .001$. A 3 (Stimulus Length: short vs. medium vs. long) univariate ANOVA was performed on both indexes. Analysis on the *d'* and *c* showed no significant main effect of stimuli length (Fs < .99, ps > .37). Hence, no differences in accuracy or response bias occurred across conditions.

2.2.4. Confidence

Participants were moderately confident about their SO judgments ($M = 4.16$, $SD = 1.36$). No significant main effects or interactions were found (Fs < 2.66, ps > .08).

2.3. Discussion

Study 1 suggested that stimuli length matters for SO judgments measured on a Kinsey-like scale but not on a SO binary choice. Overall, participants tended to be accurate in judging the sexual orientation of heterosexual but not of gay speakers, as the straight categorization bias would predict[15]. Indeed, correct answers for heterosexual, but not gay speakers, exceeded the

chance level and correct discriminability was poor when SO was assessed on a binary choice. On the Kinsey-like scale, ratings for both heterosexual and gay speakers were below the scale midpoint and toward the 'heterosexual pole' of the scale. However, on this measure, participants differentiated between gay and heterosexual speakers with the former being rated as slightly less heterosexual than the latter. Hence, a differentiation in relative terms of the two groups of speakers was observed[19, 20, 33], that does not necessarily imply absolute accuracy. Moreover, such differentiation became less relevant when participants listened to longer audio stimuli than when they had to form an impression from a short (single word) or medium (single sentence) stimuli. One possibility for this unexpected result is that longer stimuli involving rich information may provide multiple cues that make listeners think more carefully about their SO judgments and engage less in stereotyping that guide auditory gaydar judgments (see [54] for cognitive styles in gaydar judgments). Our short stimuli involved a single word with specific consonants (/s/) that have been found to be linked to perceived SO[25]. In short stimuli such information may be more salient than in longer stimuli in which vowels, consonants, and prosody are available. Stimuli length did not affect gender typicality or confidence in SO judgments.

3. Study 2

Study 2 aimed to extend our initial findings and to overcome some limitations. First of all, Study 1 referred only to male speakers and to a very small voice sample. Hence, in Study 2 we doubled the number of speakers. Previous studies have mostly focused on male speakers and rarely compared accuracy for the two sexes (for a critique see [39]). An exception is a recent study by Kachel et al. (2020) that considered both male and female targets when examining gaydar judgments on a binary choice. In this study, participants were exposed to targets' voices or faces alone, or the two combined cues. Results showed that, for both genders, accuracy was poorer

when voice was the only cue available. In terms of gender differences, no difference between male and female targets emerged when SO judgments were made from voice alone. However, this study assessed gaydar judgments and accuracy only on a binary choice. Hence, Study 2 extended this work by examining SO judgments on both binary choice and Kinsey-like scale measures. Moreover, since in Study 1 the medium stimulus did not include the same consonants of the short stimulus condition since the word 'starts' was missing in the sentence, we fixed this issue in Study 2.

3.1. Method

3.1.1. Participants

160 participants were recruited on Prolific Academic and rewarded £1.30. We excluded 10 participants who did not identify as heterosexual and one who did not consent to the use of their data. The final sample consisted of 149 (78 men, 71 women; M_{age} = 39.27, *SD* = 13.46) heterosexual English native speakers. They were all from the United Kingdom and Ireland. The majority were White (96%, *n* = 143), had a higher education qualification (75.5%, *n* = 111), and report no religion (40.8%, *n* = 60) (Table 2). Participants political orientation was balanced (left-wing: 41.6%, n = 62; center: 20.1%, n = 30; right-wing: 33.6%, n = 50; 4.7%, n = 7 did not reply[b]). Participants reported to know few gay men and lesbian women (*M* = 3.81, *SD* = 1.78) and reported to be quite unfamiliar with the gay community (*M* = 2.48, *SD* = .89).

A GPower[49] sensitivity analysis with power (1 - β) set at 0.80 and α = 05, two measurements and 6 groups indicated that the sample size allowed to detect a small effect size of .15.

[b] Participant's average self-reported political orientation (*M* = 3.80, *SD* = 1.93) and political ideology (*M* = 5.26, *SD* = .89) indicated that they were more liberal.

3.1.2. Audio stimuli

Audio recordings of 20 (10 self-identified heterosexual and 10 self-identified gay) male speakers and 20 female speakers (10 self-identified heterosexual and 10 self-identified lesbian) were obtained following the same procedure as in Study 1. Ten male speakers were the same as in Study 1. The speakers were all British with English as the first language and with no age differences between the same gender groups of speakers (male speakers: $M_{\text{heterosexual}} = 29.70$, $SD = 9.94$ vs. $M_{\text{gay}} = 27.80$, $SD = 9.60$, $t(18) = .43$, $p = .67$; female speakers: $M_{\text{heterosexual}} = 32.70$, $SD = 5.83$ vs. $M_{\text{lesbian}} = 29.30$, $SD = 6.02$, $t(18) = 1.28$, $p = .22$).

The sentences selected as stimuli were the same as in Study 1. However, in this case, the medium condition of one sentence was modified to "The English course starts on Monday" so that all the stimuli included the word used in the short stimulus condition.

3.1.3. Procedure and Materials

The procedure of the study was the same as in Study 1 but with the following exceptions. Participants listened to either 20 (10 heterosexual and 10 gay) male speakers or 20 (10 heterosexual and 10 lesbian) female speakers in one of the three stimulus (short, medium, long) condition. As in Study 1, after listening to the speaker participants completed the gender typicality, SO Kinsey-like scale, SO binary choice, and confidence measures. Also, they completed the same political ideology ($\alpha = .61$) and contact/familiarity ($\alpha = .80$) scales and reported the same demographics as in Study 1. On average the study lasted 12 minutes.

3.2. Results

A 2 (Speaker SO: heterosexual vs. gay) x 2 (Speaker Gender: male vs. female) x 3 (Stimulus Length: short vs. medium vs. long) repeated measures ANOVA was conducted on Likert scales. The first variable was within-participants and the others were between-participants

factors. Pairwise comparisons (Bonferroni correction) were performed in case of significant interactions.

3.2.1. Gender typicality

Ratings were averaged for the male and female heterosexual and LG speakers, with the female speakers' ratings being inverted to create gender typicality scores. The higher the score, the more gender typicality attributed to the speakers. LG speakers (M = 5.16, SD = .84) were rated significantly less gender typical than heterosexual speakers (M = 5.34, SD = .80), $F(1, 143) = 22.38$, $p < .001$, $\eta_p^2 = .14$. Also, female speakers (M = 5.45, SD = .77) were overall perceived as more gender typical than male speakers (M = 5.05, SD = .74), $F(1, 143) = 1.03$, $p = .001$, $\eta_p^2 = .07$. These effects were qualified by a significant interaction between Speaker SO and Speaker Gender, $F(1, 143) = 15.28$, $p < .001$, $\eta_p^2 = .10$; and by a three-way interaction between Speaker SO, Speaker Gender, and Stimulus Length, $F(2, 143) = 6.97$, $p = .001$, $\eta_p^2 = .09$. As shown in Table 2, for the male speakers, heterosexual speakers were perceived as more gender typical in the long stimulus condition ($p = .009$) but not in the other conditions (ps > .10). On the contrary, heterosexual female speakers were perceived as more gender typical than lesbian speakers in all conditions (ps < .03).

3.2.2. Sexual Orientation – Kinsey-like Scale

Ratings were averaged for the LG and heterosexual speakers separately. The higher the rating, the more likely the speakers were perceived as LG. A significant main effect of Speaker SO indicated that ratings were significantly higher for LG speakers (M = 3.26, SD = .67) than heterosexual speakers (M = 3.11, SD = .71), $F(1, 143) = 14.72$, $p < .001$, $\eta_p^2 = 0.09$. Additional t-tests against the scale midpoint showed that both heterosexual, $t(148) = -15.31$, $p < .001$, and LG speakers, $t(148) = -13.60$, $p < .001$, were rated on the heterosexual pole of the scale.

A significant three-way interaction between Speaker SO, Speaker Gender, and Stimulus Length, $F(2, 143) = 3.29$, $p = .040$, $\eta_p^2 = 0.04$, was found. As shown in Table 3, for male speakers, higher ratings for gay than heterosexual speakers occurred in the long stimulus condition ($p = .02$) but not in the other conditions (ps > .43). On the contrary, higher ratings for lesbian than heterosexual female speakers occurred in the short ($p = .002$) and medium ($p = .01$) but not in the long condition ($p = .43$). No other significant effects were found (Fs < 1.99, ps > .16).

3.2.3. Sexual orientation – Binary choice

Overall, the percentage of correct SO answers was higher for heterosexual ($M = 78.45$, $SD = 18.04$) than LG speakers ($M = 31.27$, $SD = 23.43$). Accuracy tended to be lower for heterosexual men ($M = 75.67$, $SD = 18.51$) than heterosexual women ($M = 81.20$, $SD = 17.23$), $t(147) = -1.88$, $p = .06$, while it was lower for lesbian speakers ($M = 25.87$, $SD = 22.90$) than for gay men ($M = 36.75$, $SD = 22.82$; $t(147) = -2.91$, $p = .004$). t-tests against the chance level (50%) indicated that correct answers for both heterosexual male, $t(73) = 11.93$, $p < .001$, and female speakers, $t(74) = 15.67$, $p < .001$, was above the chance level. On the contrary, correct answers for gay, $t(73) = -4.99$, $p < .001$, and lesbian speakers, $t(74) = -9.12$, $p < .001$, were below the chance level.

We then performed a signal detection analysis, considering Speaker gender differences. Overall, d' indicated a poor accuracy for male speakers ($M = .33$, $SD = 1.37$; t-test against zero: $t(73) = 2.09$, $p = .04$) and female speakers ($M = .23$, $SD = 1.49$; t-test against zero: $t(74) = 1.33$, $p = .18$). Analyses on d' showed that the interaction between speaker gender and stimuli length was not significant, $F(1, 143) = 1.83$, $p = .16$, $\eta_p^2 = .02$.

Overall, participants showed a response bias favoring a heterosexual answer for both male (M = 1.28, SD = 1.33, t-test against zero: $t(73) = 8.27$, $p < .001$) and female speakers (M = 3.23, SD = 1.50, t-test against zero: $t(74) = 9.69$, $p < .001$). No significant effects emerged on c index suggesting no difference in the response bias across conditions (Fs < 2.94, ps > .09).

3.2.4. Confidence

A significant main effect of Speaker gender showed that participants were more confident when judging the SO of male (M = 4.21, SD = 1.14) than female speakers (M = 3.25, SD = 1.55), $F(1, 143) = 17.91$, $p < .001$, $\eta_p^2 = .11$. No other significant effects or interactions emerged (Fs < 2.91, ps > .09).

3.3. Discussion

Study 2 showed once again that SO judgments were overall inaccurate for LG speakers. Indeed, accuracy was poor both on a binary choice and Kinsey-like scale, and a response bias indicating a straight categorization bias emerged. Looking at judgments on the Kinsey-like scale across speaker gender, we found that ratings for male speakers were affected by stimuli length. When long stimuli were used, gay male speakers were perceived as less gender typical and less heterosexual on a Kinsey-like scale than heterosexual male speakers. This result was not in line with Study 1's findings but suggested that heterosexual participants exposed to longer stimuli tended to differentiate gay and heterosexual male speakers in relative terms, even if such differentiation did not suggest an overall accuracy. A different pattern of results emerged for female speakers. Relative differences emerged on SO Kinsey-like scale in the short and medium, but not in the long stimuli condition. Hence, these perceived differences between lesbian and heterosexual speakers disappeared when long stimuli where involved. Gender typicality was always higher for heterosexual than lesbian speakers regardless of stimuli length condition.

For both male and female speakers, accuracy coming from categorization on a SO binary choice was not affected by stimuli length.

Interestingly, participants reported higher confidence for male than female speakers. This is in line with research suggesting that heterosexuals believe voices to be more informative of men than women' SO[55].

4. Study 3

The first two studies suggested that SO judgments on a Kinsey-like scale can be influenced by the type of stimuli. In such studies we operationalized stimulus length in terms of exposure duration (mimicking studies on visual gaydar[48]) and amount of meaningful lexical information (word, single sentence, two sentences). However, previous work[34] has shown that a progressive increase of vocal information from minimal and meaningless information (i.e., a single phoneme) to a meaningful information (word) can affect accuracy. In real life, when listening to a speaker, we are exposed incrementally to vocal and linguistic information that together form a meaningful message. Hence, it is important to examine whether exposure to sequential words that form a sentence (e.g., article: 'the' vs. article + first word: 'the dog' vs. full sentence: 'the dog runs in the park') affects SO judgments. This would also provide us an indication of which moment in the sentence processing SO judgments may change. Hence, Study 3 relied on a different stimulus length manipulation consisting in the sequential presentation and the length was determined by the word number forming the sentence.

4.1. Method

4.1.1. Participants

146 participants were recruited on Prolific Academic and rewarded £1.50. We excluded 6 participants who did not identify as heterosexual, two who did not consent to the use of their data

and one who provided rating for only two out of twenty speakers. The final sample consisted of 137 (49 men, 88 women; M_{age} = 39.27, *SD* = 13.46) heterosexual English native speakers. They were all British. The majority of the participants was White (89%, *n* = 122), had a higher education qualification (62%, *n* = 85), and report no religion (48%, *n* = 66). Participants' political orientation was mostly left (left-wing: 65%, n = 89; center: 24.1%, n = 33; right-wing: 10.3%, n = 14; 0.7%, n = 1 did not reply[c]). Participants reported to know few gay men and lesbian women (*M* = 4.41, *SD* = 1.67) and reported not to be quite very familiar with the gay community (*M* = 2.91, *SD* = 1.63).

A GPower[49] sensitivity analysis with power (1 - β) set at .80 and α = .05, two measurements and 6 groups indicated that the sample size (n = 137) allowed to detect a small effect size of .16.

4.1.2. Audio stimuli

Speakers were the same used in Study 2. We used the two sentences (i.e., "the dog ran in the park" and "the English course starts on Monday") involved in the long stimulus of the two previous studies since they both consisted of 6 words and had a similar structure. We manipulated stimulus length as the sequential exposure to linguistic information coming from one sentence. In the short condition, participants heard only the first word of the sentence (article: "The", less than half of a second). In the medium condition, they heard two words that (article + word: "The dog"/ "The English", around a second) and in the long condition, they heard the full sentence ("The dog ran in the park" or "The English course starts on Monday.", 2-3 seconds).

[c] Participant's average self-reported political orientation (*M* = 2.72, *SD* = 1.43) and political ideology (*M* = 5.49, *SD* = .71) indicated that they were liberal.

Procedure and Materials

The procedure and the dependent variables were the same as in Study 2. After listening to each speaker, participants completed the gender typicality, SO Kinsey-like scale, SO binary choice, and confidence measures. Ratings were averaged to create scores as in Study 2. At the end, they completed the political ideology (α = .59) and contact/familiarity (α = .82) scales and reported their demographics. On average the study lasted 9 minutes.

4.2. Results

A 2 (Speaker SO: heterosexual vs. gay) x 2 (Speaker Gender: male vs. female) x 3 (Stimulus Length: short vs. medium vs. long) repeated measures ANOVA was conducted on Likert scales. The first variable was within-participants, and the others were between-participants factors. Pairwise comparisons (Bonferroni correction) were performed in case of significant interactions.

4.2.1. Gender typicality

Results showed a significant main effect of speaker SO, $F(1, 131) = 17.97$, $p < .001$, $\eta_p^2 = .12$, that was qualified by speaker gender, $F(1, 131) = 6.15$, $p = .01$, $\eta_p^2 = .04$. Pairwise comparisons showed no speaker SO difference for male speakers ($M_{Gay} = 5.20$, $SD = .96$ vs $M_{Heterosexual} = 5.28$, $SD = .88$; $p = .21$) while lesbian speakers ($M = 5.09$, $SD = .95$) were perceived as less gender typical than heterosexual female speakers ($M = 5.40$, $SD = .79$; $p < .001$). No other significant main effects or interactions were found (Fs < 2.32, ps >.10).

4.2.2. Sexual Orientation – Kinsey-like Scale

Analysis showed a significant main effect of speaker SO, $F(1, 131) = 8.33$, $p = .005$, $\eta_p^2 = .06$. Overall, LG speakers ($M = 3.39$, $SD = .74$) were rated as less heterosexual than the heterosexual speakers ($M = 3.25$, $SD = .72$). Also, t-test against the scale midpoint confirm that

both heterosexual, $t(136) = -12.17$, $p < .001$, and LG speakers, $t(136) = -9.54$, $p < .001$, were rated on the heterosexual pole of the scale.

The other significant main effects or interactions failed to reach standard level of significance (Fs < 2.75, ps >.06). Because relvant for our research, we further examine a non-significant trend involving an interaction between speaker SO and condition emerged, $F(1, 131) = 2.75$, $p = .06$, $\eta_p^2 = .04$. Pairwise comparisons showed that participants rated LG speakers ($M = 3.50$, $SD = .61$) higher on SO than the heterosexual speakers ($M = 3.18$, $SD = .66$; $p = .001$) in the medium stimulus condition, while no SO difference emerged in the short ($M_{LG} = 3.43$, $SD = .64$ vs $M_{Heterosexual} = 3.32$, $SD = .64$; $p = .26$) and long ($M_{LG} = 3.25$, $SD = .89$ vs $M_{Heterosexual} = 3.25$, $SD = .72$; $p = .73$) stimulus conditions.

4.2.3. Sexual Orientation – Binary Choice

Overall, the percentage of correct SO answers was higher for heterosexual ($M = 73.75$, $SD = 18.69$) than LG speakers ($M = 29.48$, $SD = 21.22$). Accuracy was similar for heterosexual men ($M = 75.79$, $SD = 19.65$) and heterosexual women ($M = 71.64$, $SD = 17.54$; $t(134) = 1.47$, $p = .20$), as well as for gay men ($M = 26.57$, $SD = 19.02$) and lesbian women ($M = 32.53$, $SD = 23.05$; $t(135) = -1.65$, $p = .10$). Further t-tests against the chance level (50%) indicated that correct answers for both heterosexual male, $t(68) = 10.90$, $p < .001$, and heterosexual female speakers, $t(66) = 10.10$, $p < .001$, was above the chance level. On the contrary, correct answers for gay, $t(69) = -10.31$, $p < .001$, and lesbian speakers, $t(66) = -6.20$, $p < .001$, were below the chance level.

We then performed a signal detection analysis. Overall, d' indicated a poor accuracy for male speakers ($M = .08$, $SD = 1.73$) and female speakers ($M = .09$, $SD = 1.61$). Analyses on d'

considering speaker gender and stimulus length showed no significant main effects or interactions (Fs < 1.99, ps > .14).

Overall, participants showed a response bias favoring a heterosexual answer for both male (M = 1.38, SD = 1.54, t-test against zero: $t(73) = 8.27$, $p < .001$) and female speakers (M = 1.06, SD = 1.45, t-test against zero: $t(74) = 9.69$, $p < .001$). No significant speaker gender or stimulus length effects emerged on c index suggesting no difference in the response bias across conditions (Fs < 1.65, ps > .20).

4.2.4. Confidence

A significant main effect of Stimulus Length, $F(2, 130) = 4.78$, $p = .01$, $\eta_p^2 = .07$, indicated that participants were less confident about their SO judgments in the short (M = 2.95, SD = 1.31) than in the medium (M = 3.78, SD = 1.48; $p = .07$) and long stimulus condition (M = 3.92, SD = 1.32; $p = .003$), with no difference between these last two conditions (p =.24). No other significant main effects or interactions were found (Fs < 3.65, ps >.06).

4.3. Discussion

Study 3 showed that, even when stimulus length was manipulated as the sequential exposure to words in a sentence, auditory gaydar accuracy was poor. Indeed, participants were generally inaccurate when categorizing LG speakers on a binary choice and a response bias indicating a tendency to categorize speakers as heterosexual (i.e., straight categorization bias) emerged. An overall relative difference between LG and heterosexual speakers emerged when judgments were provided on a Kinsey-like scale, but both SO ratings were on the ‘heterosexual’ pole of the Kinsey-like scale.

Stimulus length played a minor role in this study, and only when SO was assessed on a Kinsey-like scale. The marginally significant interaction suggested that LG speakers tended to be

perceived as less heterosexual than heterosexual speakers if the stimulus was medium, namely an article and a word, than when the length was short (just an article) or long (a sentence). Finally, participants reported to be less confident when the stimulus was short. This is not surprising considering that the short stimulus consisted of only few phonemes (i.e., 'the') and lasted less than half second.

5. General Discussion

This research examined gaydar accuracy when different SO measures and types of stimuli were employed. In line with some previous research[10, 14, 20, 22], across two studies, we found that heterosexual listeners were generally inaccurate in judging SO of both male and female speakers. When SO judgments were assessed on a binary choice, accuracy was poor and a response bias toward the 'heterosexual' choice always occurred. On the Kinsey-like scale, ratings for both LG and heterosexual speakers were toward the 'heterosexual' pole and never exceeded the scale midpoint. These results suggest a tendency to perceive the speakers as heterosexual as the straight categorization bias[16] would predict. This speaks to the current debate on how we define accuracy[39] and whether LG ratings must fall on the 'gay/lesbian' end of the Kinsey-like scale for gaydar to be described as accurate[14]. This implies that the way accuracy is defined goes hand in hand with SO operationalization and results interpretation[37,39].

Importantly, this was the first research that directly compared SO judgments on a binary choice and a Kinsey-like scale. While accuracy on binary choice was consistently poor and clearly highlighted listeners' tendency to incorrectly categorize gay speakers as heterosexual, ratings on a Kinsey-like scale showed statistical differences on SO ratings for LG and heterosexual speakers across all the studies. Hence, what we observed in our studies was a differentiation in relative terms about how heterosexual and LG speakers' SO was perceived.

Indeed, LG speakers were perceived as 'less heterosexuals' than heterosexual speakers, possibly due to a subgroup of speakers who particularly sounded LG as observed in previous work [13, 20]. However, these different ratings did not represent an index of 'absolute' accuracy as they never exceeded the scale midpoint. These results are important as they speak to previous gaydar studies that interpreted differences in ratings as 'accurate' gaydar. By assessing SO judgments on both measures, and interpreting below midpoint ratings as inaccurate, we showed here that inaccuracy co-occur on binary and Kinsey-like scale assessments. Still, the relative differentiation occurring on the Kinsey-like scale ratings provide us relevant information. Indeed, it suggests that listeners perceived SO as a continuum as it happens for other social categories[43] and that, even if they categorize some LG speakers as heterosexual on a binary choice, they may still perceive them as less heterosexual, namely less prototypical or less similar to the reference category (heterosexuality). This informs us about a potential underlining process leading to SO categorization that can have social consequences. Since deviation from heterosexuality is usually perceived in negative terms, speakers who sounds less heterosexual, regardless of whether they are categorized as gay, can face stigmatization[55].

Our manipulation, stimuli length, did sometimes influence gaydar judgments but in an inconsistent way across studies, and only when SO was assessed on a Kinsey-like scale. In Study 1 a relative difference in SO ratings from gay and heterosexual male speakers emerged in the short and medium length stimulus conditions. In Study 2, participants rated gay and heterosexual male speakers' SO differently only in the long stimuli condition. For female speakers, SO ratings based on Kinsey-like scale were higher for lesbian than heterosexual women only in the short and medium length stimulus conditions. In Study 3, no speaker gender differences emerged. Also, although not reaching conventional levels of significance, results suggested that

participants in the medium stimulus condition tended to differentiate LG speakers from heterosexual speakers. Explaining such differences is difficult. To advance some possible explanations we looked at similarities across the studies. Except for male speakers in Study 2, a general tendency to see relative differences in ratings for LG and heterosexual speakers emerged when stimuli involved one or two words. This effect seems to suggest that listeners engage in a quick categorization that let them differentiate speakers' SO when stimuli are short and involve little vocal information. The longer the information, the less evident such differentiation seemed to be (with the exception of ratings for male speakers in Study 2). Possibly, longer stimuli allow the listeners to revise their initial categorization because additional (vocal) information is available. This should be directly tested in a future study in which listeners assess the speakers' SO sequentially after continuously listening to their voices. This would allow to examine whether SO ratings are revised during the listening process and in which exact moment.

Looking at the differences across studies, the more inconsistent findings due to stimulus length emerged for male speakers. Stimuli length influenced heterosexual listeners judging speakers' SO on a Kinsey-like scale, showing a more inconsistent pattern across studies for male speakers. While the differences between SO ratings for male speakers between study 1 and 2 could be explained by the larger and potentially more diverse voice sample of Study 2, the different pattern of result in Study 3 could be related to the different types of stimuli. Indeed, the medium stimulus length in Study 3 comprised of two words representing an intermediate condition between the short and medium stimulus conditions of Study 1 and 2 and involved words with different consonants and vowels that may have influenced the SO perception (see [34]).

No difference between stimuli occurred on gaydar judgments measured on binary choices, namely when two discrete and pre-defined categories were presented. This suggests that

there are specific features of the stimuli that influence the degree to which LG and heterosexual speakers are perceived as relatively different from each other, but what these features are remain unclear. In contrast to studies on visual gaydar [5, 48], we are yet to prove that richer information and longer exposure to stimuli can increase gaydar accuracy or that shorter stimuli may trigger faster and more polarized SO judgments[34] (see also [48] for visual cues and time). Our findings suggest that there are other features influencing SO judgments that go beyond the mere duration or type of linguistic information provided. Future studies will need to look at single phonetic components involved in each stimuli, as in Tracy and colleague's (2016), and further examine the role of linguistic content.

Finally, LG speakers were overall perceived as less gender typical than heterosexual speakers of the same gender in relative terms. Indeed, overall, speakers were always perceived as gender typical, but LG speakers were perceived a little bit less gender typical. There were instances that type of stimuli influenced gender typicality, such as for gay male speakers for whom a lower gender typicality in the long stimuli condition was found in Study 2. Future studies should address this point by examining how such changes in gender typicality due to stimuli could influence SO ratings. Moreover, recent work has shown that not only the listeners' perceived gender typicality but also the speakers' self-reported gender typicality matter[22]. Hence, future studies should further examine the role of speaker's self-perception on gaydar judgments. Indeed, speakers' motives and self-perception of sounding gay can influence their behavior and elicit voice modulation [21, 56, 57].

5.1. Limitations and Future Directions

Several limitations need to be considered. First, in our studies we have operationalized length in terms of exposure duration (mimicking studies on visual gaydar[48]), amount of linguistic

information (word, single sentence, two sentences) or progressive increase of vocal and lexical information (article, article + word, full sentence). We also relied on stimuli coming from two sentences, hence limiting the type of content (e.g.., consonant, vowels) examined here. Future studies should involve testing a variety of linguistic stimuli. This will allow researchers to understand which is the information that is more likely to elicit SO differentiation on a Kinsey-like scale. Also, previous work has shown that longer sentences improved intelligibility compared to words sliced from sentences[58]. This may affect listeners' perception of speakers. Moreover, we did not examine prosody in relation to SO judgments. Prosody varies between words and sentences and hence likely varied across our stimuli. Prosody may be particularly important when comparing reading and spontaneous speech. So far, auditory gaydar research has mostly examined SO judgments made after listening to speakers reading words, sentences, or texts (but see [56] for an exception). In spontaneous speech, words may be uttered differently as function of the connected speech[59]. Hence, specific words would sound differently in reading and spontaneous speech[60]. Future research should take this into consideration and examine whether prosody and intelligibility change across stimuli, as well as whether differences between reading and spontaneous speaking play a role on listeners' SO judgments.

Another issue is related to the Kinsey-like scale that we used. We relied on the same scale used in previous work for replicability reasons. However, this Kinsey-like scale varying from 'exclusively heterosexual' to 'exclusively gay/lesbian' is simplistic: it still considers only two main SO categories and thus does not account for other SO (e.g., bisexuality[39]). Hence, it could potentially be interpreted as a measure of prototypicality, namely whether speakers are perceived as more or less prototypical of heterosexual speakers (for voice prototypicality see [61]). Future research should expand our research and consider multiple SO categories and a more

inclusive Kinsey-like measure that better capture the concept of SO fluidity. For instance, Kachel et al. (2020) measured self-perceived sexuality on a Kinsey-like scale but also on a multidimensional measure of sexual orientation involving items concerning romantic feelings and sexual attraction. If used to assess listeners' perceptions, such kind of measures may allow us to understand judgments and expectations that goes beyond a mere categorization of speakers as gay or heterosexual.

Finally, our findings refer to a limited number of speakers. We cannot generalize our results to all LG and heterosexual speakers. Significant variations occur within groups of LG and heterosexual speakers[22], and speakers' self-perceptions[21,22] and concealment motives (e.g., coming out) influencce SO judgments[56,57]. Future research should consider larger samples of speakers and how the way they self-perceive may affect gaydar judgments. Our research adds to the few studies comparing SO judgments for male and female speakers. As in previous work[22], on binary choices we did not find speaker gender differences, but some differences emerged on the Kinsey-like scale. Future studies should carry on studies looking at speaker gender and why differences may occur. Similarly, attention should be given to listeners. In our studies, participants were similar in terms of demographics, but we did not assess listeners' individual differences. Since individuals' endorsement of sexual prejudice[62] or motivations to avoid labelling someone as LG[63] are associated with a stronger straight categorization bias, future research should consider these variables.

6. Conclusion

This research aimed to examine the impact that SO measure and various aspects of stimulus can have on auditory gaydar research. In much prior research (Table 1), conclusions have been made about a unitary gaydar on the basis of a single-gender speaker group, a single type of stimuli, and

a binary choice or Kinsey-like scale. This research indicates the importance of considering how we conceptualize and operationalize SO judgments and gaydar accuracy. Overall, this research showed that auditory gaydar is generally inaccurate and that SO judgments depend on the measures used to assess them. Indeed, judgments on binary choices led to overall poor accuracy that was not influenced by type of stimuli. Judgments on Kinsey-like scale showed a relative differentiation of LG speakers as less heterosexual than heterosexual speakers, but such judgments were influenced, albeit and in an inconsistent way, by the type of stimuli. Contributing to previous work and current debates[37, 39], this research suggests scholars to carefully think about the stimuli they use when conducting auditory gaydar studies and to explain how they conceptualize (i.e., relative vs. absolute differentiation) and assess gaydar accuracy.

References

1. Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within-and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*(6), 806-819. https://doi.org/10.1177%2F0146167206286627
2. McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PloS one*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779
3. Belin, P., Boehme, B., & McAleer, P. (2017). The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PloS one*, *12*(10), e0185651. https://doi.org/10.1371/journal.pone.0185651
4. Fasoli, F., Maass, A., & Sulpizio, S. (2016). Communication of the "invisible": Disclosing and inferring sexual orientation through visual and vocal cues. In *Advances in Intergroup Communication* (pp. 193–208). Peter Lang Publications.
5. Rule, N. O. (2017). Perceptions of sexual orientation from minimal cues. *Archive of Sexual Behaviour, 46*(1), 129-139. https://doi.org/10.1007/s10508-016-0779-2
6. Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: evidence from obvious and ambiguous groups. *Current Directions in Psychological Science*, *26*(3), 231-236. https://doi.org/10.1177%2F0963721417697970
7. Epstein, R., McKinney, P., Fox, S., & Garcia, C. (2012). Support for a fluid-continuum model of sexual orientation: A large-scale internet study. *Journal of Homosexuality*, *59*(10), 1356-1381. https://doi.org/10.1080/00918369.2012.724634

8. Kinsey, A. C., Pomeroy, W. R., & Martin, C. E. (2003). Sexual behavior in the human male. *American Journal of Public Health*, *93*(6), 894-898.
9. Barton, B. (2015). How like perceives like: Gay people on "gaydar". *Journal of Homosexuality, 62*(12), 1615-1637.
10. Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society, 32*(3), 329-350. https://doi.org/10.1017/S0047404503323024
11. 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*(2), 202-240. https://doi.org/10.1016/j.wocn.2005.05.003
12. Rieger, G., Linsenmeier, J. A., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior, 39*(1), 124-140. https://doi.org/10.1007/s10508-008-9405-2
13. 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*(7), 1560-1578. https://doi.org/10.1044/2018_JSLHR-S-17-0125
14. 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*(1), 184-206. https://doi.org/10.1177%2F0023830919828201
15. Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. *The Journal of the Acoustical Society of America*, *141*(6), 4793-4809. https://doi.org/10.1121/1.4988684

16. Lick, D. J., & Johnson, K. L. (2016). Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments. *Journal of Personality and Social Psychology, 110*(6), 801–817. DOI: https://doi.org/10.1037/pspa0000052
17. Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, *11*(1), 83-96. https://doi.org/10.1111%2Fj.1471-6402.1987.tb00776.x
18. Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin*, *36*(10), 1318-1331. https://doi.org/10.1177%2F0146167210378755
19. Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and speech, 50*(1), 125-142. https://doi.org/10.1177/00238309070500010601
20. 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*(7), e0128882. https://doi.org/10.1371/journal.pone.0128882
21. Fasoli, F., Hegarty, P., Maass, A., & Antonio, R. (2018). Who wants to sound straight? Sexual majority and minority stereotypes, beliefs and desires about auditory gaydar. *Personality and Individual Differences, 130*, 59-64. https://doi.org/10.1016/j.paid.2018.03.046
22. Kachel, S., Steffens, M. C., Preuß, S., & Simpson, A. P. (2020). Gender (conformity) matters: Cross-dimensional and cross-modal associations in sexual orientation

perception. *Journal of Language and Social Psychology*, *39*(1), 40-66. https://doi.org/10.1177%2F0261927X19883902

23. Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2018). Investigating the common set of acoustic parameters in sexual orientation groups: A voice averaging approach. *PloS one, 13*(12), e0208686. https://doi.org/10.1371/journal.pone.0208686
24. Crist, S. (1997). Duration of onset consonants in gay male stereotyped speech. *University of Pennsylvania Working Papers in Linguistics, 4*(3), 4. Available at: https://repository.upenn.edu/pwpl/vol4/iss3/4
25. 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*(1), 198-212. https://doi.org/10.1016/j.wocn.2011.10.002
26. Moonwoman-Baird, B. (1997). Toward the study of lesbian speech. In A. Livia & K. Hall (Eds.), *Queerly phrased* (pp. 202–213). Oxford University Press
27. Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1-74. https://doi.org/10.1016/S0065-2601(08)60317-2
28. Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*(5), 399-427. https://doi.org/10.1016/0022-1031(87)90038-2

29. Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not so fast: The (not-quite-complete) dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin*, *36*(2), 264-277. https://doi.org/10.1177%2F0146167209354519

30. Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PloS one*, *13*(10), e0204991. https://doi.org/10.1371/journal.pone.0204991

31. Ferdenzi, C., Patel, S., Mehu-Blantar, I., Khidasheli, M., Sander, D., & Delplanque, S. (2013). Voice attractiveness: Influence of stimulus duration and type. *Behavior Research Methods*, *45*(2), 405-413. https://doi.org/10.3758/s13428-012-0275-0

32. Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review*, *17*(1), 72-86. https://doi.org/10.1177%2F1088868312461308

32. Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech, 69*(1), 30-57. http://dx.doi.org/10.2307/455948

33. Valentova, J. V., & Havlíček, J. (2013). Perceived sexual orientation based on vocal and facial stimuli is linked to self-rated sexual orientation in Czech men. *PloS one, 8*(12), e82417. https://doi.org/10.1371/journal.pone.0082417

34. Tracy, E. C., Bainter, S. A., & Satariano, N. P. (2015). Judgments of self-identified gay and heterosexual male speakers: Which phonemes are most salient in determining sexual orientation? *Journal of Phonetics, 52*, 13-25. https://doi.org/10.1016/j.wocn.2015.04.001

35. Campbell-Kibler, K. (2007). Accent,(ING), and the social logic of listener perceptions. *American Speech*, *82*(1), 32-64. https://doi.org/10.1215/00031283-2007-002

36. Campbell-Kibler, K. (2011). Intersecting variables and perceived sexual orientation in men. *American Speech*, *86*(1), 52-68. https://doi.org/10.1215/00031283-1277510

37. Gelman, A., Mattson, G., & Simpson, D. (2018). Gaydar and the fallacy of decontextualized measurement. *Sociological Science, 5*: 270-280. DOI: 10.15195/v5.a12

38. Dunham, Y., & Olson, K. R. (2016). Beyond discrete categories: Studying multiracial, intersex, and transgender children will strengthen basic developmental science. *Journal of Cognition and Development*, *17*(4), 642-665. https://doi.org/10.1080/15248372.2016.1195388

39. Miller, A. E. (2018). Searching for gaydar: Blind spots in the study of sexual orientation perception. *Psychology & Sexuality, 9*(3), 188-203. https://doi.org/10.1080/19419899.2018.1468353

40. Master, A., Markman, E. M., & Dweck, C. S. (2012). Thinking in categories or along a continuum: Consequences for children's social judgments. *Child Development*, *83*(4), 1145-1163. https://doi.org/10.1111/j.1467-8624.2012.01774.x

41. Foroni, F., & Rothbart, M. (2011). Category boundaries and category labels: When does a category name influence the perceived similarity of category members?. *Social Cognition*, *29*(5), 547-576. https://doi.org/10.1521/soco.2011.29.5.547

42. Maddox, K. B., & Dukes, K. N. (2008). Social categorization and beyond: How facial features impact social judgment. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 205–233). Guilford Publications.

43. Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, *98*(6), 3080-3095. https://doi.org/10.1121/1.413832

44. Morandini, J., Dacosta, L., & Dar-Nimrod, I. (2021) Heterosexuals exposed to continuous or fluid theories of sexual orientation are more likely to embrace non-exclusive orientations. *Scientific Reports.*

45. Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93-120. https://www.annualreviews.org/doi/abs/10.1146/annurev.psych.51.1.93

46. Scheffey, K. L., Ogden, S. N., & Dichter, M. E. (2019). "The idea of categorizing makes me feel uncomfortable": University student perspectives on sexual orientation and gender identity labeling in the healthcare setting. *Archives of Sexual Behavior*, *48*(5), 1555-1562. https://doi.org/10.1007/s10508-019-1401-1

47. Brewer, G., & Lyons, M. (2016). Discrimination of sexual orientation: Accuracy and confidence. *Personality and Individual Differences*, *90*, 260-264. https://doi.org/10.1016/j.paid.2015.11.019

48. Rule, N. O. & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology, 44*(4), 1100-1105. https://doi.org/10.1016/j.jesp.2007.12.001

49. Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1-11. https://doi.org/10.3758/BF03203630

50. Crist, S. (1997). Duration of onset consonants in gay male stereotyped speech. *University of Pennsylvania Working Papers in Linguistics, 4,* 53–70.

51. Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica*, *50*(1), 35-48. https://doi.org/10.1159/000021447

52. Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America*, *116*(4), 1905-1908. https://doi.org/10.1121/1.1788729

53. Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149. https://doi.org/10.3758/BF03207704

54. Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2013). The politics of gaydar: Ideological differences in the use of gendered cues in categorizing sexual orientation. *Journal of personality and social psychology*, *104*(3), 520-541. https://psycnet.apa.org/doi/10.1037/a0031187

55. 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. https://doi.org/10.1111/bjso.12442

56. Daniele, M., Fasoli, F., Antonio, R., Sulpizio, S., & Maass, A. (2020). Gay voice: Stable marker of sexual orientation or flexible communication device?. *Archives of Sexual Behavior*, *49*(7), 2585-2600. https://doi.org/10.1007/s10508-020-01771-2

57. Daniele, M., Fasoli, F., & Johnson, S. (2020). Coming out as a factor of voice change: From sounding straight to sounding gay. *Psicologia Sociale*, *15*(1), 33-52. https://www.rivisteweb.it/doi/10.1482/96293

58. Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*(1), 79-84. https://doi.org/10.1016/S0022-5371(64)80062-1

59. Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *Journal of Speech and Hearing Disorders*, *55*(2), 360-363. https://doi.org/10.1044/jshd.5502.360

60. Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*(2), 163-169. https://doi.org/10.1016/0167-6393(91)90039-V

61. Dragojevic, M., Giles, H., Beck, A. C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385-405. https://doi.org/10.1080/03637751.2017.1322213

62. Rule, N. O., Tskhay, K. O., Brambilla, M., Riva, P., Andrzejewski, S. A., & Krendl, A. C. (2015). The relationship between anti-gay prejudice and the categorization of sexual orientation. *Personality and Individual Differences*, *77*, 74-80. https://doi.org/10.1016/j.paid.2014.12.038

63. 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*(6), 1266-1289. https://psycnet.apa.org/doi/10.1037/pspi0000232

64. Vasilovsky, A. T. (2018). Aesthetic as genetic: The epistemological violence of gaydar research. *Theory & Psychology, 28*(3), 298-318. https://doi.org/10.1177/0959354318764826

65. Carahaly, L. N. (2000). *Listener accuracy in identifying the sexual orientation of male and female speakers* (Doctoral dissertation, The Ohio State University).

66. Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 533-554. https://www.jstor.org/stable/20108037

67. Piccolo, F. (2008). Perceived sexual orientation and attitudes towards sounding gay or straight. *University of Pennsylvania Working Papers in Linguistics, 14(2),* 130-138. https://repository.upenn.edu/pwpl/vol14/iss2/16/

68. Cuddy, S. (2019). *Can women" sound gay"?: A sociophonetic study of/s/and pitch of gay and straight British-English speaking women* (Doctoral dissertation, University of York). http://etheses.whiterose.ac.uk/27430/1/Cuddy%202019.pdf

69. Carraro, L., Castelli, L., & Macchiella, C. (2011). The automatic conservative: Ideology-based attentional asymmetries in the processing of valenced information. *PLoS One, 6*(11), e26456-e26456. https://doi.org/10.1371/journal.pone.0026456

Table 1. Overview of Studies Testing Voice-based Perception of Sexual Orientation. Studies are presented in order of year publication. Subscript refers to reference number.

Paper	Speakers' Gender	Number of Speakers	Type of Stimulus	Sexual Orientation Measurement
Moonwomon-Baird (1986)[26]	Female	6 lesbian women, 6 heterosexual women	Spontaneous speech/conversation (30s)	Binary categories
Gaudio (1994)[32]	Male	4 gay men, 4 heterosexual men	Paragraph	Kinsey-like scale
Linville (1998[510]	Male	5 gay men, 4 heterosexual men	Monologue	Binary categories
Carahaly (2000)[65]	Male and Female	10 gay men, 10 heterosexual men, 10 lesbian women, 10 heterosexual women	Spontaneous speech/conversation & paragraph combined (32s)	Binary categories
Smyth, Jacobs, & Rogers (2003)[10]	Male	17 gay men, 8 heterosexual men	Paragraph (30s)	Binary categories
Pierrhumnert, Bent, Munson, & Bradlow, & Bailey (2004)[52]	Male and Female	29 gay men, 26 heterosexual men, 16 lesbian women, 16 bisexual women, 16 heterosexual women	Multiple sentences	Kinsey-like scale
Munson, McDonald, DeBoe, & White (2006) – Study 2[11]	Male and Female	11 gay men, 11 heterosexual men, 11 lesbian/bisexual women, 11 heterosexual women	Sequence of 3 words	Kinsey-like scale
Levon (2007)[66]	Male	1 gay-sounding,1 heterosexual-sounding	Paragraph	Kinsey-like scale
Munson (2007)[19]	Male and Female (same speakers as Munson et al., 2006)	11 gay men, 11 heterosexual men, 11 lesbian/bisexual women, 11 heterosexual women	Words	Kinsey-like scale
Piccolo (2008)[67]	Male	6 gay men, 6 heterosexual men	Paragraph (30/45s)	Kinsey-like scale

Rieger, Linsenmeier, Gygax, Garcia, & Bailey (2010)[12]	Male and Female	25 gay men, 25 heterosexual men, 23 lesbian women, 22 heterosexual women	Sentence/spontaneous speech (6/10s)	Kinsey-like scale
Mack & Munson (2012)[25]	Male	8 gay-sounding,8 heterosexual-sounding	Single word	Kinsey-like scale
Valentova & Havlíček (2013)[33]	Male	27 gay men, 34 heterosexual men	Paragraph (20s)	
Sulpizio, Fasoli, Maass, Paladino, Vespignani, Eyssel, & Bentler (2015)[20]	Male	16 gay men, 16 heterosexual men	Single sentence (5s)	Kinsey-like scale (Study 1) - Binary categories (Study 2 and 3)
Tracy, Bainter, & Satariano (2015)[34]	Male	18 gay men, 18 heterosexual men	Single word (Study 1) and single phoneme (Study 2)	Kinsey-like scale
Kachel, Simpson, & Steffens (2018)[13]	Male	25 gay men, 26 heterosexual men	Single sentence	Kinsey-like scale
Kachel, Radtke, Skuk, Zaske, Simpson, & Steffens (2018)[23]	Male and Female	5 gay men, 5 bisexual men, 5 heterosexual men, 5 lesbian women, 5 bisexual women, 5 heterosexual women	Single sentence	Kinsey-like scale
Cuddy (2019)[68]	Female	Heterosexual woman (perception study)	Single sentences	Likert scale
Sulpizio, Fasoli, Antonio, Eyssel, Paladino, & Diehl (2020)[14]	Female	14 lesbian women, 15 heterosexual women	3 sentences (15 seconds)	Kinsey-like scale
Kachel, Steffens, Preuß, & Simpson (2020)[22]	Male and Female	18 gay men, 18 heterosexual men, 18 lesbian women, 18 heterosexual women	Single sentence	Binary categories
Daniele, Fasoli, Antonio, Sulpizio, & Maass (2020)[56]	Male	10 gay men and 10 heterosexual men (Study 1) 7 gay men and 7 heterosexual men (Study 2)	Sentence/spontaneous speech (Study 1: 15 seconds and Study 2: 30 seconds)	Kinsey-like scale

Daniele, Fasoli, & Johnson (2020)[57]	Male	1 gay man and 1 heterosexual man (Study 1) 3 gay men and 3 heterosexual men (Study 2)	Sentence/spontaneous speech (30 seconds)	Kinsey-like scale

Table 2

Means (Standard Deviations) for Gender Typicality across Speaker SO, Speaker Gender, and Stimulus Length (Study 2).

		Short Stimulus	**Medium Stimulus**	**Long Stimulus**
Male	Gay	5.10 (.82)	5.22 (.65)	4.77 (.87)
	Heterosexual	4.95 (.90)	5.20 (.69)	5.03 (.69)
Female	Lesbian	5.04 (.83)	5.54 (1.00)	5.28 (.77)
	Heterosexual	5.55 (.67)	5.82 (.85)	5.50 (.65)

Table 3

Means (Standard Deviations) for Kinsey-like scale ratings across Speaker SO, Speaker Gender, and Stimulus Length (Study 2).

Speaker		**Short Stimulus**	**Medium Stimulus**	**Long Stimulus**
Male	Gay	3.22 (.72)	3.10 (.65)	3.25 (.67)
	Heterosexual	3.24 (.82)	3.02 (.65)	3.01 (.64)
Female	Lesbian	3.48 (.57)	3.24 (.75)	3.26 (.63)
	Heterosexual	3.18 (.58)	2.99 (.80)	3.19 (.76)

Daniel Painter has recently graduated in Psychology at the University of Surrey. Daniel is interested in research on minority groups, social inequality, and social policies.

Fabio Fasoli is a Lecturer in Social Psychology at the University of Surrey. His research focuses on voice self-perception, auditory gaydar, stereotyping and discrimination, and derogatory language.

Simone Sulpizio is Assistant Professor at the Department of Psychology, University of Milano-Bicocca (Italy). His two main research interests are the neurocognitive mechanisms underlying word reading, and the interplay between language processing and social and emotional factors.