Perceived Vocal Congruence Varies Across Gender Identities

Chiara De Livio (chiara.delivio@uniroma1.it)
Department of Psychology, Sapienza University of Rome, Rome, Italy

Claudia Mazzuca (claudia.mazzuca@uniroma1.it)
Department of Dynamic and Clinical Psychology and Health Studies,
Sapienza University of Rome, Rome, Italy

Chiara Fini (chiara.fini@uniroma1.it)
Department of Dynamic and Clinical Psychology and Health Studies,
Sapienza University of Rome, Rome, Italy

Anna M. Borghi (anna.borghi@uniroma1.it)
Department of Dynamic and Clinical Psychology and Health Studies,
Sapienza University of Rome, Rome, Italy
Institute of Cognitive Sciences and Technologies, Italian National Research Council, Rome, Italy

Abstract

This study investigates vocal congruence across populations with different gender identities. Forty-four participants completed a self-voice perception task in three conditions (Silent Reading, Reading Aloud, and Listening to their recorded voice) after reading gender-stereotyped priming texts. Our findings show that transgender and gender non-conforming participants experience lower vocal congruence listening to their outer voice compared to cisgender participants, and they perceive their inner voice as more congruent to the self. Results confirm the role of interoceptive sensibility on general voice congruence perception, suggesting that it varies across gender identities. Further research is needed to deepen the relationship between inner experience and voice perception and to disentangle the reciprocal relationship between self-identity and self-voice perception.

Keywords: vocal congruence; interoceptive sensibility; gender identity; embodied cognition

Introduction

Body awareness relies on the ability to integrate multisensory information coming both from the inside and the outside (Tsakiris, Jiménez & Costantini 2011). To form a unitary sense of self, information must be perceived as congruent. One common experience of incongruence is the mismatch between the voice perceived while speaking and the actual voice that other people hear, or our own recorded voice from a voice message—a phenomenon widely known as *voice confrontation* (Holzman & Rousey, 1966). In fact, speech production requires the multimodal integration of many information sources (auditory feedback, proprioceptive, tactile, and barometric; Ito, Tiede, & Ostry, 2009; Haggard & de Boer, 2014; Kent, 2024). A recent survey comprising 1,5k U.S. participants showed that almost 40% of respondents reported discontent with their voice, without correlation with voice disorders (Naunheim, Puka & Huston, 2023). Discontent with one's voice has been associated with sociodemographic variables like gender, age, education level and health conditions, although evidence is mixed (Gregory et al., 2016; Naunheim, Puka & Huston, 2023; Roy, Merrill, Gray, Smith, 2005; Verdonck-de Leeuw & Mahieu, 2004).

While the incongruence is driven by the physical transformations our voice goes through between the production and the hearing phases (Pörschmann, 2000; Orepic et al., 2023; Stenfelt, 2016), the sense of discomfort relates more specifically to the tight connection between voice and identity. Indeed, voice has been defined as an *auditory face* (Belin, Fecteau & Bedard, 2004). Rather than being just a vehicle for language, it is a salient marker of individuality (Crow, van Mersbergen & Payne, 2021). It allows to reveal or conceal personal traits like provenience, age, or gender (Kreiman & Sidtis, 2011; Sidtis and Kreiman, 2012), and to flexibly adapt in social contexts (Chong, Choi, & Lee, 2022; Guldner, Nees & McGettigan, 2020).

Not only we use the voice to interact with others, but we also interact with ourselves through inner verbalization. Inner speech is a form of internal language that many people report as a significant feature of their subjective experience (Alderson-Day & Fernyhough, 2015; Fernyhough & Borghi, 2023; Loevenbruck et al., 2018; Perrone-Bertolotti et al., 2014) and it spans over different levels of intentionality, ranging from mind-wandering to dialogic forms, including variations like silent reading (Perrone-Bertolotti et al., 2012; Vilhauer, 2016; Yao et al., 2011).

Being so central in participating to the bodily and psychological self, the voice can be a source of intense discomfort when it does not align with our desired identity. In this regard, Crow and colleagues (2021, p.1) defined vocal congruence as "the extent to which one's voice is in alignment, or congruent, with one's sense of self" and designed a self-report measure, the Vocal Congruence Scale (VCS), to investigate the degree of individual identification with the voice and its associated beliefs. By using a heart-beat detection task and an interoceptive confidence task to assess interoceptive awareness, the authors showed that the higher the vocal congruence is, the higher the interoceptive awareness. Moreover, while the VCS moderately correlated

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

with interoceptive awareness and confidence it did not, instead, correlate with interoceptive accuracy—showing that metacognitive judgments are more relevant than interoceptive accuracy for self-voice perception.

As of today, these findings can only be generalized to the cisgender population (i.e., people whose birth sex aligns with gender identity). However, exploring how changes of identity and voice mutually influence each other could be crucial, especially for the transgender and gender non-conforming population (TGNC; i.e., people whose gender identity varies from assumptions based on their birth sex, e.g., transgender or nonbinary individuals; APA, 2015). The voice is one of the most commonly mentioned bodily sources of distress by TGNC people, with speech therapists often reporting that these clients feel their outer voice does not match with their "inner" and "true" voice (van de Grift., 2016a; van de Grift et al., 2016b; Ziltzer et al., 2023). Also, voice alteration represents a central goal in the gender-affirmation process (James et al., 2016; Kennedy & Thibeault, 2020). In fact, among the multitude of gender communication cues, voice plays a key role in gender categorization, possibly exposing TGNC individuals to unwanted episodes of birth sex disclosure (Chang & Yung, 2021), hence impacting their quality of life and the possibility of being perceived as their affirmed gender in their social community (dos Santos Oliveira et al., 2024; Hancock, 2017).

Culturally inherited stereotypes about what is considered feminine or masculine are often enacted through language (Charlesworth et al., 2021; Jones et al., 2020; Lewis & Lupian, 2020; Sczesny, Formanowicz & Moser, 2016; Skewes, Fine & Haslam, 2018), which serves as a cognitive cue to reinforce culturally normative self-construals (Kemmelmeier & Cheng 2004). Studies with priming procedures showed that gender cues can impact on both explicit and implicit associations and attitudes of stereotyped group members (Pesciarelli, Scorolli & Cacciari, 2019; Steele & Ambady, 2006), and cause self-perception and social behaviour to become more congruent to gender stereotypes (Hundhammer & Mussweiler, 2012). Investigating the effects of gendered stereotypes and discrimination is crucial because of the effects of this long-term exposure to stress within the TGNC community (Kallitsounaki, & Williams 2023, Mazzoli et al., 2022; Mirabella et al., 2020). Breaking the binary representation of gender identity, TGNC individuals face continuous discrimination, and inequality and social stigma in all aspects of their lives (Carmel & Erickson-Schroth, 2016; Connolly et al., 2016; Drabish & Theeke, 2022; Pinna et al., 2022; Romani et al., 2021; Russell & Fish., 2016; Testa et al., 2015; Truszczynski, Singh & Hansen, 2022). However, little is known about the impact of gendered stereotypes conveyed by language on vocal congruence and, to the best of our knowledge, no study has systematically investigated vocal congruence perception both in cisgender and gender non-conforming individuals.

To sum up, voice can be considered as a bodily feature contributing to the sense of identity. It is the medium through which we interact with ourselves and with others, and it is enriched by cultural meanings. Among these, aspects related to gender identity are particularly relevant for the self-voice perception of specific populations like TGNC individuals, whose distress related to the perceived incongruence between their voice and gender identity might be further exacerbated by experiences of discrimination.

The Present Study

In this study we investigate vocal congruence across populations with different gender identities (cisgender and gender non-conforming) using a behavioural self-voice perception task. We build on existing studies of vocal congruence by introducing an additional condition—the perception of inner voice congruence—to previously studied scenarios, such as reading aloud and listening to one's own recorded voice (Crow et al., 2021; Welch & Helou, 2022). Importantly, participants are presented with excerpts of texts conveying either gender stereotypical content (feminine vs. masculine) or gender-neutral content. Overall, we hypothesise differences in voice perception between the cisgender and the gender non-conforming populations, with the latter experiencing lower vocal congruence across the three conditions, reporting lower congruency in the conditions where they are asked to focus their attention on the outer voice (i.e., Reading Aloud and Listening) compared to the Silent Reading. We further hypothesise that vocal congruence would be affected by linguistic gender-stereotyped primes, especially for the more stereotyped gender identity group (i.e., gender non-conforming). Although a stricter test of the priming effect should have targeted specific gender identities within each group paired with the relevant linguistic stereotype, the scarcity of the sample prevented us from deepening this aspect. Nonetheless, we still expected the gendered semantic content of texts would specifically impact on TGNC participants vocal congruence perception, because it might draw participants' attention to the general gender conceptual dimension. Finally, we sought to further address the role of interoception on vocal congruence. Previous findings suggest higher interoceptive awareness is associated with higher vocal congruence (Crow et al., 2021). Here, we predict that belief associated with body perception, i.e., interoceptive sensibility, might partially explain lower congruency of self-voice perception, specifically for gender non-conforming participants.

Method

Ethical approval was provided by the Ethics Committee of the Sapienza University of Rome (Ethical Approval 0000856).

Participants

A total of 44 Italian native speakers took part to the study. Recruitment was carried out via word-of-mouth and through

the involvement of an LGBTQIA+ rights local association and a hospital service dedicated to gender-affirmation processes. Trainer vocalists and individuals who were at advanced stages of the gender-affirmation process and had undergone voice training interventions were not eligible. Participants were asked to provide general information about birth sex, gender identity, and sexual orientation. Most participants (57%, n = 25) were assigned female at birth, whereas 41% (n = 18) were assigned male at birth and one participant was assigned as intersex. Two groups were identified based on gender identity: transgender and gender non-conforming participants (n = 22) and participants who identify as cisgender (n = 22). In the cisgender group, 55% of the participants identify as cisgender woman (n = 12) and the 45% identify as cisgender man (n = 10). In the TGNC group, 32% of the participants identify as transgender man (n = 7), 23% identify as transgender woman (n = 5), 23% identify as non-binary (n = 5). Four participants (18%) provided their gender identity using the "other" response option The two groups were comparable in terms of age, $t(36.33) = -0.087$, $p = .930$ (M cisgender = 27.59; SD = 8.04; M TGNC = 27.31; SD = 12.21), with an average age of the overall sample of 27.45 years (SD = 10.22; Age-range = 18 – 66). In the cisgender group, 36% (n = 8) participants reported having a high school diploma, 36% (n = 8) a master's degree, and 18% (n = 4) a PhD. In the TGNC group, 14% (n = 3) completed lower high school, 50% (n = 11) participants reported having a high school diploma, 36% (n = 3) have a bachelor's degree, and 18% (n = 4) have a PhD.

Procedure

The experimental procedure was divided in three parts that participants completed in two sessions. In the first session, participants first completed the voice perception task to assess levels of vocal congruence, and then in the second part of the first session, they were asked to provide ratings on different words on various semantic dimensions. Finally, they gave demographic information. In the second session, participants completed questionnaires investigating interoceptive sensibility, emotion regulation, alexithymia, gender identity, and gender discrimination. Questionnaires were administered in a second session of the experimental procedure, after two days from the completion of the first session, to avoid fatiguing effects and carryover effects of the voice perception task. For space reasons, here we will focus on the first part of session one (i.e., the vocal congruence task) and on session two (see Figure 1).

The experimental procedure was implemented with Qualtrics, using an on-line questionnaire divided into sections that participants filled in a fixed order. Participants seated in front of a computer in a silent room. In the first block participants were presented with four texts (Neuter text 1, Masculine text, Neuter text 2, Feminine text) in a fixed order, to avoid a possible carryover effect of the gender manipulation. For each text, participants had to complete a Silent Reading, a Reading Aloud and a Listening task. In the Silent Reading condition participants were instructed to silently read the text, with no pressure about time. In the Reading Aloud condition, they were instructed to read the same text aloud and to register themselves while reading by using the online questionnaire's interface. In the Listening condition, participants were instructed to listen to their recorded voice. The order of the three conditions was kept fixed to avoid carryover effects of both Reading Aloud and Listening conditions on the Silent Reading condition (see Discussion for possible limitations). Participants were reassured they could stop listening to the registrations at any moment. After each condition participants were presented with an Italian version of the Vocal Congruence Scale (Crow et al., 2021; see Materials). At the end of each text block, participants were asked to rate the perceived congruence between the inner voice heard during the Silent Reading and the voice heard during the Listening condition (Inner-Outer Voice Congruence Ratings: 1 = *completely incongruent*; 5 = *completely congruent*). We added this explicit question to better understand the priming role of the gender-stereotyped text on the experience of vocal congruence. After the experimental task, participants were presented with the questionnaires (see Materials). At the end of the procedure, participants gave general demographic information about birth sex, gender identity, sexual orientation, birth country, language, and educational level. Birth sex was investigated using a multiple-choice question with three alternatives: *male*, *female*, and *intersex*. Gender identity was investigated by asking participants to indicate their gender identity through a multiple-choice question with the following alternatives: *agender*, *cisgender man*, *cisgender woman*, *non-binary*, *transgender man*, *transgender woman*, and *other* (this response option was followed by an open-ended response box). To investigate sexual orientation, we implemented an open-ended response format.

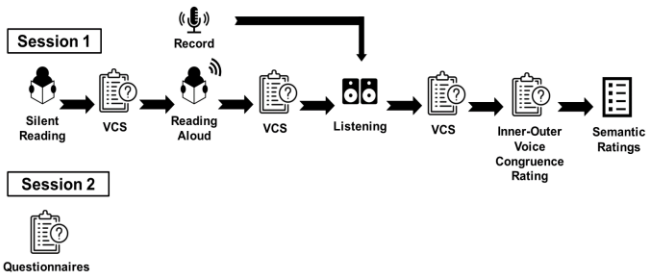

Figure 1: Schematic representation of the experimental procedure implemented for each type of text (feminine, masculine, neuter).

Materials

Stimuli Four texts were chosen based on the presence or absence of a clear gender stereotypic representation of feminine and masculine gender identities. The two gender-neuter texts were taken from "Come vivevano i Greci" (Paoli, 1957) and from "Un mare di silenzio" (Rava, 2012). The two gendered texts were taken from "L'abbecedario degli

Stereotipi di Genere" (Priulla & Sammartino, 2020). All the texts had the same length and were easily understandable since they are taken from education and dissemination books.

The questionnaires presented in the second session were: (i) The Multidimensional Assessment of Interoceptive Awareness (Mehling et al., 2012), (ii) The Emotion Regulation Questionnaire (Gross & John, 2003), (iii) The Toronto Alexithymia Scale (Bagby, Parker & Taylor, 1994), and (iv) The Multi-Gender Identity Questionnaire (Joel et al., 2014). In addition to these, participants of the TGNC group completed the Gender Minority Stress and Resilience Measure (Testa et al., 2015). For the purposes of this paper, here we only focus on MAIA.

Vocal Congruence Scale (VCS) 10-item self-report (0 = *strongly disagree*; 4 = *strongly agree*) questionnaire designed to measure vocal congruence. The questionnaire was back-translated and the scoring adapted for this study (1 = *strongly disagree*; 5 = *strongly agree*).

Inner-Outer Voice Congruence Rating To measure the effect of the type of text manipulation we constructed a 5-points Likert scale (1 = *completely incongruent*; 5 = *completely congruent*) rating that we administered after each text to test the congruency between the inner voice perceived during the Silent Reading and the voice heard during the Listening condition.

Multidimensional Assessment of Interoceptive Awareness (MAIA) 32-items self-report questionnaire assessing interoceptive awareness across eight constructs: (i) *Noticing*: awareness of uncomfortable, comfortable, or neutral body sensations; (ii) *Non-Distracting*: avoiding distraction to cope with bodily discomfort; (iii) *Not-Worrying*: tendency not to experience emotional distress about bodily discomfort; (iv) *Attention Regulation*: ability to sustain and control attention on the body; (v) *Emotional Awareness*: internal process involving the ability to attribute specific physical sensations to physiological manifestations of emotions; (vi) *Self-Regulation*: ability to regulate distress by attention to body sensations; (vii) *Body-Listening*: tendency to actively attend to body signals for insights; (viii) *Trust*: experience of one's body as safe and trustworthy.

Data Analysis

Data pre-processing, analysis and visualization were carried out with R (R Core Team, 2022) and RStudio (Posit Team, 2023). Data were first inspected to identify incomplete responses. After removing one participant who did not complete all the questionnaires, pre-processing was carried out using "dplyr" R's package (Wickham et al., 2022).

As a preliminary step, we run separate *t*-tests to check whether the two neuter texts and the two gendered texts differed one from the other in terms of their impact on vocal congruence. We found no difference between the two neuter texts, $t(1319) = -0.574$, $p = .565$ and we found no difference between the gendered texts, $t(1319) = -0.139$, $p = .888$.

To address our first research question, linear mixed-effects models were fitted using "lme4" R's package (Bates et al., 2015). As we were interested in investigating the role of gender stereotypes in affecting the perception of vocal congruence in different conditions as a function of different gender identity experiences, we fit a linear mixed-effects model with total scores of the Vocal Congruence Scale as dependent variable; Group (gender non-conforming vs. cisgender); Text (neuter vs. feminine vs. masculine); Condition (Silent Reading vs. Reading Aloud vs. Listening); a two ways interaction between Group and Condition and a two-ways interaction between Group and Text as fixed factors; participants as random intercepts. Post-hoc contrasts were carried out with the "emmeans" R's package (Lenth, 2023) using Tukey's adjustment for multiple comparisons.

To assess the relationship between our Inner-Outer Voice Congruence Rating measure of vocal congruence and the VCS scale we computed their Spearman correlation. We found a significant positive correlation between the two measures, $r(44) = 0.66$, $p < .0001$. To investigate whether the gender priming had an effect on vocal congruence scores, we fitted a cumulative link mixed model with the "ordinal" R's package (Christensen, 2022) featuring scores of the Inner-Outer Voice Congruence Rating as dependent variable; Group (gender non-conforming vs. cisgender); Text (neuter vs. masculine vs. feminine), and their interaction as fixed factors; and participants as random intercepts.

Finally, to account for the potential impact of interoceptive awareness on vocal congruence, we used the main model adding the subdimensions of the MAIA questionnaire as covariates. Post-hoc contrasts were carried out with the "emtrends" function in the "emmeans" R's package (Lenth, 2023). Model comparisons were conducted using the "anova" function from the "stats" R's package (R Core Team, 2023).

Vocal congruence is affected by gender experiences

We found a significant main effect of Group, $\chi^2(1) = 3.86$, $p = .049$, a significant main effect of Condition, $\chi^2(2) = 132.62$, $p < .0001$, and a significant two-ways interaction between Group and Condition, $\chi^2(2) = 63.19$, $p < .0001$ (see Figure 2). No other main effect or interaction reached significance, all $ps > .190$.

Pairwise comparisons showed that TGNC participants gave higher scores of vocal congruence in the Silent Reading compared to the Reading Aloud ($z = 6.398$, $SE = .741$, $p < .0001$) and Listening conditions ($z = 8.000$, $SE = .741$, $p < .0001$), while there was no difference between the Reading Aloud and the Listening conditions ($z = 1.602$, $SE = .741$, $p = .078$). In the cisgender group, we found no significant differences between Silent Reading and Reading Aloud, ($z = -0.261$, $SE = .741$, $p = .933$), between Silent Reading and Listening, ($z = 0.432$, $SE = .741$, $p = .829$) and between Reading Aloud and Listening ($z = 0.693$, $SE = .741$, $p = .618$), although VCS scores tended to decrease across conditions

also in this sample, possibly due to a general effect of voice confrontation (see Figure 2).

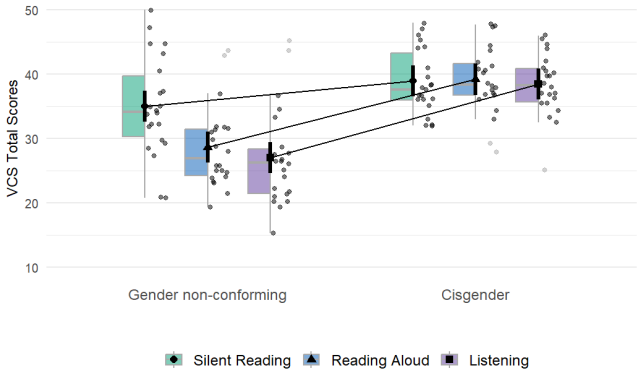

Figure 2: Distribution of predicted VCS scores across the three conditions for TGNC and cisgender participants across the three types of text. Thick black dots represent estimated marginal means and their standard errors (vertical thick black line), and grey dots represent raw data.

No effect of gendered priming on vocal congruence

We found a significant main effect of Group, χ^2 (1) = 25.11, p <.0001. No other main effect reached significance, all *ps* > .211. This highlights that—independently from the priming stereotyped text—TGNC participants experienced a lower vocal congruence between their inner voice and their recorded voice compared to cisgender participants, β = 5.821, *SE* = 1.169, *z* = 4.977, p <.0001.

Interoceptive sensibility enhances vocal congruence also depending on gender experiences

Before assessing whether levels of interoceptive sensibility differentially impact vocal congruence across groups, we first inspected the explanatory power of each MAIA's subdimension on vocal congruence overall. Since the main model (see above) showed that Text did not affect vocal congruence, we removed the Text factor from the model and compared the two resulting models. The two models did not differ, χ^2 (4) = 6.402, p = .171, so we rely on the second model and ran eight separate models adding each MAIA's subdimension. Across all the subdimensions, three had a significant main effect on the VCS scores overall.

The Not Distracting model showed a significant main effect of the covariate, χ^2 (1) = 4.96, p = .025, a significant main effect of Group, χ^2 (1) = 7.068, p =.007, a significant main effect of Condition, χ^2 (2) = 130.879, p < .0001, and a significant two-way interaction between Group and Condition, χ^2 (2) = 62.365, p < .0001.

The Not Worrying model showed a significant main effect of the covariate, χ^2 (1) = 4.319, p = .037, a significant main effect of Group, χ^2 (1) = 28.632, p < .0001, a significant main effect of Condition, χ^2 (2) = 69.409, p < .0001, and a significant two-way interaction between Group and Condition, χ^2 (2) = 62.365, p < .0001.

Finally, the Self-Regulation model showed a significant main effect of the covariate, χ^2 (1) = 8.244, p =.004, a significant main effect of Condition, χ^2 (2) = 130.879, p < .0001, and a significant two-way interaction between Group and Condition, χ^2 (2) = 62.365, p < .0001.

Once assessed the impact of the covariates overall, we turned to address their interaction with Group. Only for the Not Worrying model we found a significant interaction between Group and the covariate, χ^2 (2) = 4.770, p = .028. Simple slope analysis revealed that only the slope of TGNC group (LCL = 0.83; UCL = 4.45) was significantly different from zero as a function of the continuous predictor Not Worrying. While the cisgender group showed a higher voice congruence as compared with the gender non-conforming group, the latter showed higher voice congruence at higher levels of Not Worrying score (see Figure 3).

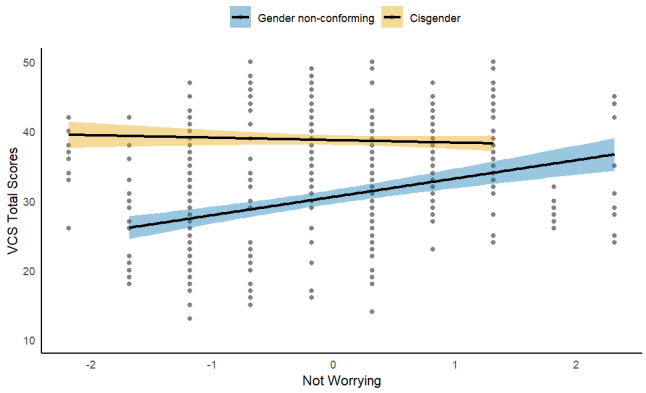

Figure 3: Interaction between Group and Not Worrying (centred) on VCS scores. Shaded coloured bands represent 95% confidence intervals.

As for Not Distracting and Self Regulation, we found that higher Not Distracting scores were associated with lower perceived vocal congruence regardless the experimental condition and of the group, and higher Self Regulation scores were associated with higher perceived vocal congruence, regardless the experimental condition and of the group.

Discussion

In this study we investigated vocal congruence across different gender identities, comparing cisgender and transgender and gender non-conforming participants (TGNC) on a self-voice perception task in three conditions: (i) Silent Reading, (ii) Reading Aloud, and (iii) Listening to their recorded voice. We predicted that TGNC participants would report lower scores of vocal congruence in conditions in which they were asked to pay attention to their outer voice, compared to cisgender participants. Also, we expected that reading a gender-stereotyped text would decrease vocal

congruence in TGNC participants. Results confirm our first prediction, with TGNC participants experiencing lower vocal congruence in both Reading Aloud and Listening conditions, but not in the Silent Reading condition, compared to cisgender participants. Interestingly, TGNC participants perceived their inner voice as more congruent to the self, compared to the outer voice. This is consolidated by the results coming from the Inner-Outer Voice Congruence Rating, with TGNC participants reporting a greater mismatch between their internally perceived voice and their recorded voice, compared to cisgender participants. While these findings highlight the voice incongruence experienced by the gender non-conforming population, an alternative explanation for the reported mismatch could be that some participants did not use inner speech, although this seems unlikely given that they were instructed to read silently. Furthermore, this would not explain the differences found between the two groups—albeit there is evidence that not all humans experience inner speech to the same degree, with specific populations possibly refraining from using it at all (e.g., Alderson-Day & Pearson, 2023). Further research on the role of inner speech in voice self-perception studies is required.

Our second prediction pertained to the role of linguistic gender stereotypes in affecting the perception of bodily-related cues (i.e., voice) in relation to gender identity. However, we found no differences across the two groups as a function of the type of text (gendered *vs*. neutral), with the gender non-conforming group always displaying higher vocal incongruence regardless of the gendering of the text.

Previous studies show a relationship between self-voice perception and interoception (Crow et al., 2021; Orepic et al., 2022; Smeltzer, Chiou & Shembel, 2023). In this study, we further addressed the role of interoceptive sensibility in voice perception within different gender identities. Here, we found that higher scores of Not Distracting (i.e., avoiding distraction to cope with bodily discomfort) corresponded to lower scores of vocal congruence, and higher scores of Self-Regulation (i.e., ability to regulate distress by attention to body sensations) corresponded to higher scores of vocal congruence, for both cisgender and TGNC participants. Interestingly, only for the gender non-conforming group, the less emotional distress with sensations of pain or discomfort (Not Worrying) the higher the perceived vocal congruence. Therefore, individuals who are more attuned to their internal sensations and use them to regulate psychological distress tend to exhibit better self-voice recognition. However, TGNC individuals show greater self-voice recognition when their interoceptive sensibility does not lead to significant distress and discomfort arising from the discrepancy between their actual and desired bodily characteristics.

To the best of our knowledge, this is the first study to directly compare vocal congruence between two balanced groups of cisgender and TGNC participants. Our findings are in line with previous literature on voice perception in TGNC individuals (Kennedy & Thibeault, 2020; van de Grift., 2016a; van de Grift et al., 2016b; Ziltzer et al., 2023), adding layers on this literature by tackling the unexplored topic of vocal congruence in this population. However, this study has some methodological limitations that are worth mentioning. First, we did not control for levels of vocal discomfort before the experimental procedure. Second, while the order of task conditions and type of texts in the self-voice perception task was fixed to avoid carryover effects, this could have led to further unexplored order effects—e.g., participants could have become progressively more acquainted with the active listening of their own voice (i.e., habituation effects), or could have noticed more distinctively possible differences between diverse instances of their voice through different conditions (i.e., context effects). In addition, the selection of linguistic stimuli in its own could have limited the effect of the gender primes on VCS scores. Finally, the difficulty in recruiting TGNC participants led to an overall small sample size that limited the generalisation of the results, and did not allow comparisons between different gender identities within the two groups, also as a function of stereotyped texts. Future research might better control for these aspects to better frame the relationship between vocal congruence, language, and gender identity.

Conclusion

This study investigates vocal congruence in two groups that differ in their gender identity. We found that transgender and gender non-conforming participants experience lower vocal congruence with their outer voice compared to cisgender participants, and they perceive their inner voice as more congruent to the self, compared to the outer voice. Aspects of interoceptive sensibility (Not Distracting and Self-Regulation) play a role in general voice congruence perception, while Not Worrying seems to be more important for transgender and gender non-conforming individuals. Further research is needed to deepen the relationship between inner experiences and voice perception, and to disentangle the reciprocal relationship between self-identity and self-voice perception.

Acknowledgments

We thank the Gay Center LGBT+ Center of Rome and the SAIFIP service (San Camillo-Forlanini Hospital Complex, Rome) for their help in the recruitment of participants. We are also grateful to Guido Giovanardi, Marta Mirabella and all the BalLab (Body, Action, and Language Lab) members for insightful comments on this project.

Funding sources. This work was supported by Sapienza's Excellent Project Grant n. RG12117A5D1EB0B3: "Concepts in interaction with others and with ourselves: abstractness in social interaction, metacognition and mind wandering - 2022-24" led by Anna M. Borghi.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychological bulletin, 141(5), 931.

Alderson-Day, B., & Pearson, A. (2023). What can neurodiversity tell us about inner speech, and vice versa? A theoretical perspective. *cortex*.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American psychologist*, *70*(9), 832-864.

Bagby RM, Parker JDA, Taylor GJ (1994) The twenty-item Toronto alexithymia scale—I. Item selection and cross-validation of the factor structure. J Psychosom Res 38:23–32.

Bates, D. M. (2010). lme4: Mixed-effects modeling with R.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences*, *8*(3), 129-135.

Carmel, T. C., & Erickson-Schroth, L. (2016). Mental health and the transgender population. Journal of psychosocial nursing and mental health services, 54(12), 44-48.

Chang, J., & Yung, K. (2021). Gender affirming voice care: A literature review. *International Journal of Head and Neck Surgery*, *12*(3), 93-97.

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. Psychological Science, 32(2), 218-240.

Chong, H. J., Choi, J. H., & Lee, S. S. (2022). Does the perception of own voice affect our behavior?. *Journal of Voice*.

Christensen, R. H. B. (2022). ordinal - Regression Models for Ordinal Data. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal.

Connolly, M. D., Zervos, M. J., Barone II, C. J., Johnson, C. C., & Joseph, C. L. (2016). The mental health of transgender youth: Advances in understanding. Journal of Adolescent Health, 59(5), 489-495.

Crow, K. M., van Mersbergen, M., & Payne, A. E. (2021). Vocal congruence: the voice and the self measured by interoceptive awareness. *Journal of Voice*, *35*(2), 324-e15.

dos Santos Oliveira, J. C., da Trindade Duarte, J. M., Simões-Zenari, M., & Nemr, K. (2024). Risk of Dysphonia, Presence of Vocal Changes, and Vocal Self-Perception in Brazilian Transgender Women. *Journal of Voice*.

Drabish, K., & Theeke, L. A. (2022). Health impact of stigma, discrimination, prejudice, and bias experienced by transgender people: a systematic review of quantitative studies. Issues in mental health nursing, 43(2), 111-118.

Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. Trends in cognitive sciences, 27(12):1180-1193.

Gregory, N. D., Chandran, S., Lurie, D., & Sataloff, R. T. (2012). Voice disorders in the elderly. Journal of Voice, 26(2), 254-258.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. Journal of personality and social psychology, 85(2), 348.

Guldner, S., Nees, F., & McGettigan, C. (2020). Vocomotor and social brain networks work together to express social traits in voices. *Cerebral Cortex*, *30*(11), 6004-6020.

Haggard, P., & de Boer, L. (2014). Oral somatosensory awareness. *Neuroscience & Biobehavioral Reviews*, *47*, 469-484.

Hancock, A. B., & Pool, S. F. (2017). Influence of listener characteristics on perceptions of sex and gender. *Journal of Language and Social Psychology*, *36*(5), 599-610.

Holzman, P. S., & Rousey, C. (1966). The voice as a percept. Journal of Personality and Social Psychology, 4(1), 79.

Hundhammer, T., & Mussweiler, T. (2012). How sex puts you in gendered shoes: Sexuality-priming leads to gender-based self-perception and behavior. Journal of Personality and Social Psychology, 103(1), 176–193. https://doi.org/10.1037/a0028121

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, *106*(4), 1245-1248.

James, S., Herman, J., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. A. (2016). The report of the 2015 US transgender survey.

Joel, D., Tarrasch, R., Berman, Z., Mukamel, M., & Ziv, E. (2014). Queering gender: Studying gender identity in 'normative'individuals. Psychology & Sexuality, 5(4), 291-321.

Jones, J. J., Amin, M. R., Kim, J., & Skiena, S. (2020). Stereotypical Gender Associations in Language Have Decreased Over Time. *Sociological Science*, *7*(1), 1–35. doi:10.15195/v7.a1

Kallitsounaki, A., & Williams, D. M. (2023). Brief Report: An Exploration of Alexithymia in Autistic and Nonautistic Transgender Adults. Autism in Adulthood.

Kemmelmeier, M., & Cheng, B. Y.-M. (2004). Language and Self-Construal Priming: A Replication and Extension in a Hong Kong Sample. Journal of Cross-Cultural Psychology, 35(6), 705-712. https://doi.org/10.1177/0022022104270112.

Kennedy, E., & Thibeault, S. L. (2020). Voice–gender incongruence and voice health information–seeking behaviors in the transgender community. American journal of speech-language pathology, 29(3), 1563-1573.

Kent, R. D. (2024). The Feel of Speech: Multisystem and Polymodal Somatosensation in Speech Production. *Journal of Speech, Language, and Hearing Research*, 1-37.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician*, *34*(4), 216-221.

Lewis, M., Lupyan, G. Gender stereotypes are reflected in the distributional structure of 25 languages. Nat Hum Behav 4, 1021–1028 (2020). https://doi.org/10.1038/s41562-020-0918-6

Lœvenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., & Dohen, M. (2018). A cognitive neuroscience view of inner language. *Inner speech: New voices*, *131*.

Mazzoli, F., Cassioli, E., Ristori, J., Castellini, G., Rossi, E., Romani, A., ... & Fisher, A. D. (2022). Appearent autistic traits in transgender people: a prospective study of the impact of gender-affirming hormonal treatment. The Journal of Sexual Medicine, 19(11), S60-S61.

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *PloS one*, *7*(11), e48230.

Mirabella, M., Giovanardi, G., Fortunato, A., Senofonte, G., Lombardo, F., Lingiardi, V., & Speranza, A. M. (2020). The body I live in. Perceptions and meanings of body dissatisfaction in young transgender adults: A qualitative study. Journal of Clinical Medicine, 9(11), 3733.

Naunheim, M. R., Puka, E., & Huston, M. N. (2023). Do You Like Your Voice? A Population-Based Survey of Voice Satisfaction and Voice Enhancement. *The Laryngoscope*.

Orepic, P., Kannape, O. A., Faivre, N., & Blanke, O. (2023). Bone conduction facilitates self-other voice discrimination. *Royal Society Open Science*, *10*(2), 221561.

Orepic, P., Park, H. D., Rognini, G., Faivre, N., & Blanke, O. (2022). Breathing affects self-other voice discrimination in a bodily state associated with somatic passivity. Psychophysiology, 59(7), e14016.

Paoli, U. E. (1957). Come vivevano i Greci. Edizioni Radio Italiana.

Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., ... & Lachaux, J. P. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. Journal of Neuroscience, 32(49), 17554-17562.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behavioural Brain Research, 261, 220–239. doi:10.1016/j.bbr.2013.12.034.

Pesciarelli, F., Scorolli, C., & Cacciari, C. (2019). Neural correlates of the implicit processing of grammatical and stereotypical gender violations: A masked and unmasked priming study. Biological Psychology, 146, 107714. doi:10.1016/j.biopsycho.2019.06.002.

Pinna, F., Paribello, P., Somaini, G., Corona, A., Ventriglio, A., Corrias, C., ... & Italian Working Group on LGBTQI Mental Health. (2022). Mental health in transgender individuals: a systematic review. International review of psychiatry, 34(3-4), 292-359.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. Acta Acustica united with Acustica, 86(6), 1038-1045.

Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Priulla, G., Banci, M., and Sammartino, G. (2020). *L'abbecedario degli stereotipi di genere*.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rava, C. (2012). Un mare di silenzio. Garzanti.

Romani, A., Mazzoli, F., Ristori, J., Cocchetti, C., Cassioli, E., Castellini, G., ... & Fisher, A. D. (2021). Psychological wellbeing and perceived social acceptance in gender diverse individuals. The journal of sexual medicine, 18(11), 1933-1944.

Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: prevalence, risk factors, and occupational impact. *The Laryngoscope*, *115*(11), 1988-1995.

Russell, S. T., & Fish, J. N. (2016). Mental health in lesbian, gay, bisexual, and transgender (LGBT) youth. Annual review of clinical psychology, 12, 465-487.

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination?. Frontiers in psychology, 7, 25.

Sidtis, D., & Kreiman, J. (2012). In the beginning was the familiar voice: Personally familiar voices in the evolutionary and contemporary biology of communication. *Integrative Psychological and Behavioral Science*, *46*, 146-159.

Skewes, L., Fine, C., & Haslam, N. (2018). Beyond Mars and Venus: The role of gender essentialism in support for gender inequality and backlash. PloS one, 13(7), e0200921.

Smeltzer, J. C., Chiou, S. H., & Shembel, A. C. (2023). Interoception, voice symptom reporting, and voice disorders. Journal of Voice.

Steele, J. R., & Ambady, N. (2006). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, *42*(4), 428–436. doi:10.1016/j.jesp.2005.06.003.

Stenfelt, S. (2016). Model predictions for bone conduction perception in the human. *Hearing Research*, *340*, 135–143. doi:10.1016/j.heares.2015.10.014

Testa, R. J., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the gender minority stress and resilience measure. Psychology of sexual orientation and gender diversity, 2(1), 65.

Truszczynski, N., Singh, A. A., & Hansen, N. (2022). The discrimination experiences and coping responses of non-binary and trans people. Journal of homosexuality, 69(4), 741-755.

Tsakiris, M., Jiménez, A. T., & Costantini, M. (2011). Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. Proceedings of the Royal Society B: Biological Sciences, 278(1717), 2470-2476.

van de Grift, T. C., Cohen-Kettenis, P. T., Elaut, E., De Cuypere, G. R. E. T. A., Richter-Appelt, H., Haraldsen, I. R., & Kreukels, B. P. (2016b). A network analysis of body

satisfaction of people with gender dysphoria. *Body image*, *17*, 184-190.

van de Grift, T. C., Cohen-Kettenis, P. T., Steensma, T. D., De Cuypere, G., Richter-Appelt, H., Haraldsen, I. R., ... & Kreukels, B. P. (2016a). Body satisfaction and physical appearance in gender dysphoria. *Archives of sexual behavior*, *45*, 575-585.

Verdonck-de Leeuw, I. M., & Mahieu, H. F. (2004). Vocal aging and the impact on daily life: a longitudinal study. Journal of voice, 18(2), 193-202.

Vilhauer, R. P. (2016). Inner reading voices: An overlooked form of inner speech. Psychosis, 8(1), 37-47.

Welch, B., & Helou, L. B. (2022). Measuring communicative congruence and communicative dysphoria in a sample of individuals without voice disorders. *Journal of Speech, Language, and Hearing Research*, *65*(9), 3420-3437.

Wickham H, François R, Henry L, Müller K (2022). _dplyr: A Grammar of Data Manipulation_. R package version 1.0.10, <https://CRAN.R-project.org/package=dplyr>.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of cognitive neuroscience*, *23*(10), 3146-3152.

Ziltzer, R. S., Lett, E., Su-Genyk, P., Chambers, T., & Moayer, R. (2023). Needs assessment of gender-affirming face, neck, and voice procedures and the role of gender dysphoria. *Otolaryngology–Head and Neck Surgery*, *169*(4), 906-916.