INTERNATIONAL JOURNAL OF TRANSGENDER HEALTH
2024, VOL. 25, NO. 4, 719–737
https://doi.org/10.1080/26895269.2023.2237009

Taylor & Francis
Taylor & Francis Group

Short-term effects of a speech feminization program for transgender women: listener perceptions, self-perception and satisfaction of the voice

Clara Leyns[a], Iris Meerschman[a], Guy T'Sjoen[b,c] and Evelien D'haeseleer[a,d]

[a]Center for Speech and Language Sciences (CESLAS), Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium; [b]Department of Endocrinology, Ghent University Hospital, Ghent, Belgium; [c]Center for Sexology and Gender, Ghent University Hospital, Ghent, Belgium; [d]Department of Otorhinolaryngology, Ghent University Hospital, Ghent, Belgium

ABSTRACT

Purpose: This study measured and compared the short-term impact of pitch elevation training (PET) and articulation-resonance training (ART) in transgender women, on self-perception, satisfaction and masculinity-femininity perceptions of listeners.***Methods:*** A randomized controlled study with cross-over design was used. Thirty transgender women were included and received fourteen weeks of speech training. All participants started with sham training (four weeks), after which they were randomly assigned to one of two groups: one group continued with PET (five weeks), followed by ART (five weeks), the second group received both trainings in opposite order. Participants were recorded four times, in between the training blocks: pre, post 1 (after sham), post 2 (after training 1) and post 3 (after training 2). Participants did a self-evaluation through the Trans Woman Voice Questionnaire (TWVQ) and visual analogues scales (VAS) concerning their self-perception and satisfaction. Two listening experiments ($n = 75$) were conducted researching the continuous masculinity-femininity rating (through a VAS) and categorical masculinity-femininity attribution.
Results and conclusions: Transgender women perceive their voices more feminine after the training and experience a positive impact on the vocal functioning and the voice-related impact on their daily life. However, a lot of the participants acknowledge that they need more speech training after ten weeks. Listeners rate the participants' voices more feminine after training, both during the continuous and categorical questions. Higher femininity scores were detected during self-perception and listener perceptions after the combination of both ART and PET, compared to the separate trainings. No order effects were detected between ART and PET, both for self-perception and listener perceptions. Defining outcome predictors is crucial in future research.

KEYWORDS
Effectiveness; perception; RCT; speech therapy; transgender voice

Introduction

People whose gender differs from the gender they were presumed to be at birth are usually referred to as "trans" or "gender diverse" people, also referred to as TGD persons (American Psychological Association, 2015; Coleman et al., 2012, 2022; Grant et al., 2011; Haas et al., 2014; Vincent, 2018). For some of these individuals, the inconsistency between their gender identity and the presumed gender at birth creates discomfort (Schneider et al., 2017). Research showed that the degree of satisfaction of this population with their own body is directly related to mental health outcomes (Kanamori & Xu, 2022; Novais Valente Junior & Mesquita de Medeiros, 2022; Villas-Bôas et al., 2021). The high prevalence of mental health problems among TGD people compared to cisgender people (Millet et al., 2017) has been found to correlate with the impact of minority stress (Tan et al., 2020; Testa et al., 2015) and social stigma. The well-being of TGD persons is formed by challenges and resources, and can be executed externally, i.e., distal processes, or internally by TGD persons themselves, i.e., proximal processes (Azul et al., 2022; Meyer, 2003). The degree of satisfaction with the body and thus their own voice could enact as a

CONTACT Clara Leyns clara.leyns@ugent.be Center for Speech and Language Sciences (CESLAS), Department of Rehabilitation Sciences, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium.

proximal resource or stressor, whereas attribution of masculinity-femininity to voice by listeners would be a distal stressor. Voice is one aspect of gender expression that a transgender person may wish to modify to affirm their gender (Oates, 2019). Voice training provided by a speech-language pathologist (SLP) can prevent negative mental health outcomes (Davies et al., 2015; Klemmer et al., 2021), especially because gender-affirming hormone treatment does not affect the voice of transgender women (T'Sjoen et al., 2019). WPATH's new eighth edition of the Standards of Care (SOC-8, Coleman et al., 2022) defined the purposes of speech therapy or voice and communication training TGD persons. This training should not only educate clients on the aspects contributing to gender perception by listeners (Leung et al., 2018), but also enable them to communicate their sense of sociocultural belonging in everyday encounters in a manner that matches the client's desired self-presentation and expression. SLPs help them develop, maintain and habituate their voice voices, vocal qualities, and communication practices that support the clients' goals in a manner that does not harm the voice production mechanism (Coleman et al., 2022).

Perception of masculinity-femininity ratings by listeners, self-perception, and satisfaction are essential aspects to investigate during and after gender-affirming voice and speech training. Besides the speaker's anatomy of the voice organ, gender-related voice characteristics are also influenced by speaker's voice use practices, listener practices, professional practices, and supra-individual socio-cultural factors (Azul & Hancock, 2020). The production of speaker's gender is considered as an ongoing, socio-culturally mediated, interactional process beyond the individual's control. In the study by Azul and Hancock (2020), reconceptualization of clinicians', speakers' and listeners' agency as bioculturally mediated capacity to act implies a lack of control over the production of voice and speaker socio-cultural positionings in vocal encounters. Researching a client's perspective on how they perceive their voice and how listeners perceive their voice, might help the SLP to better counsel about the unpredictability of voice production and speakers' socio-cultural positioning in daily life conversations (Azul & Hancock, 2020). The authors' theoretical model (i.e., ASSEMBLE approach) about gender affirming speech therapy practices addressed the complexity of working toward client goals that relate to listener perceptions and attributions of gender in communication. The review by Leyns et al. (2021) researched the lister perceptions, self-perception, satisfaction of the voice and acoustic outcomes of speech feminization therapy for transgender women and included 14 studies.

Concerning listening perceptions by others, several studies found significant increases in femininity and decreases in masculinity in transgender women after speech training (Bralley et al., 1978; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Kaye et al., 1993; Leyns et al., 2021; Quinn & Swain, 2018). Due to poor inter-rater reliability, mixed results were reported by Carew et al. (2007). Considering the ASSEMBLE view on gender or masculinity-femininity attribution, a poor inter-rater reliability could actually be expected because listeners might have different constructs and experiences about the attribution of gender to a voice. These results were mainly collected with Likert scales and VAS. Binary attribution has been a way to investigate the listener perception of speaker gender as well, with a higher number of female attribution post training in the study by Gelfer and Tice (2013). For many transgender clients, being perceived by the environment according to the experienced gender is very important. Including a categorical attribution experiment using a categorical scale is useful as it can reflect to a certain degree how they are perceived in daily life, although this also depends on the reference framework of the listener. It might be important for some binary transgender people who wish to be perceived binary. During real-life conversations, speakers tend to be assigned to a certain gender category such as female/male. The degree of gender (e.g., continuous scale from very feminine to very masculine), will not be reflected in conventional forms of address and referral. However, considering the ASSEMBLE view on gender or masculinity-femininity attributions during listening experiments, different

constructions of gender in response to one voice are to be expected. Transparency around the listeners' characteristics is important to interpret the gender or masculinity-femininity attributions.

Including patient-reported outcome measures (PROMs) in research enables to truly capture whether the transgender individual experiences a decreased discomfort with their voice after the training. The review by Leyns et al. (2021) found PROMs which were collected through interviews, VAS, Likert scales and standardized questionnaires. The studies reported increased satisfaction (Bralley et al., 1978; Carew et al., 2007; Dacakis, 2000; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Quinn & Swain, 2018), more self-confidence (Bralley et al., 1978; Hancock & Helenius, 2012), higher self-perception of femininity (Carew et al., 2007; Quinn & Swain, 2018; Söderpalm et al., 2004), lower scores on the TSEQ (the Self Evaluation of Voice Questionnaire, (Davies & Goldberg, 2006)), TVQ (Transsexual Voice Questionnaire, (Dacakis et al., 2013)) or Likert scale, i.e., lower impact on the voice-related functioning in daily life or limitations concerning communication (Hancock & Helenius, 2012; Mészáros et al., 2005; Quinn & Swain, 2018). Quinn et al. (2022) concluded that transgender participant self-ratings of femininity-masculinity were correlated with listener ratings, but this correlation was not strong. On the other hand, the self-ratings of femininity of the transgender participants had a consistently stronger relationship with their self-rated vocal satisfaction.

Several aspects should be considered when conducting a listening experiment. Firstly, the composition of the listening panel can have an impact on the ratings. For example, Hope and Lilley (2022) reported that transgender and non-binary listeners were more likely to use the entire expanse of the rating scales (femininity, masculinity and "other" scale) and showed systematic categorization of gender-neutral voices as non-binary. Additionally, Carew et al. (2007) advised to use SLPs in future research who are experienced in the area of voice to carry out perceptual ratings due to a poor inter-rater reliability with naïve listeners. However, it is still important to collect data from naïve listeners as well, as these people might perceive the voices of the participants differently and might give insights about daily life perception. Secondly, the tool to collect these ratings differed across previous studies (Leung et al., 2018, 2021). Nominal attribution scales were primarily used to rate speaker gender and had response options such as "male" or "female" (Hancock et al., 2014; Honorof & Whalen, 2010; Lass et al., 1976; Smith & Patterson, 2005). Ordinal scales included either Likert scales or visual analogue scales (VAS), with various options at the extremes of the scales, and were used for ratings of vocal femininity–masculinity (Addington, 1968; Cartei et al., 2014; Gelfer & Schofield, 2000; Hardy et al., 2020). Perception of masculinity-femininity of the voice can be apprehended as categorical or on a continuum. As in current literature these two different conceptions exist, both masculinity-femininity attribution (categorical, M/F/X) and masculinity-femininity ratings (continuous, VAS) were used in the current study.

Although previous research has shown that speech therapy can lead to a more feminine voice, both acoustically and perceptually, there is still a need for well-designed prospective RCTs, including larger sample sizes, well described therapy contents, multidimensional voice assessments, and blinded investigators (Leyns et al., 2021). The aim of the current study was to measure and compare the short-term impact of both pitch elevation training (PET) and articulation-resonance training (ART) in transgender women on self-perception, satisfaction and masculinity-femininity perceptions of listeners. It was hypothesized that both training programs would induce a higher perception of femininity, as these programs contain several elements that contribute to perception of a feminine speaker (Leung et al., 2018; Leyns et al., 2021).

Methods

This research project was completed according to the Declaration of Helsinki and approved by the Ethics Committee of the Ghent University Hospital with the following registration number: B670201941335. This trial has been registered on ClinicalTrials.gov, a resource provided by the U.S. National Library of Medicine. Its unique

identifier is NCT04708600. A written informed consent was signed by each participant.

Researcher positionings

Leyns C. (MS, SLP, PhD) is a cisgender woman (pronouns she/her) with 7 years of experience in voice and 4 years of experience in transgender voice. She has been researching the effectiveness of speech therapy for transgender women during her PhD. She is a member of a multidisciplinary gender team and the World and European Professional Association of Transgender Health since the start of her PhD in 2019. Her role on this project was conducting the RCT as the SLP who provided the therapy sessions, data analysis and manuscript writing.

Meerschman I. (MS, SLP, PhD) is a cisgender woman (pronouns she/her) with 9 years of experience in voice diagnostics and therapy. She has researched the effectiveness of voice therapy with multiple RCTs. She was involved in this project during statistical data analysis and manuscript writing.

T'Sjoen G. (MD, PhD) is a cisgender man (pronouns he/him) with 22 years of experience in transgender health research. He has, as endocrinologist, acted as a principal investigator in the endocrine part of the ENIGI study (European Network for the Investigation of Gender Incongruence), he is a past president of EPATH (European Professional Association for Transgender Health) and is member of a multidisciplinary gender team. His role on this project was patient recruitment and manuscript writing.

D'haeseleer E. (MS, SLP, PhD) is a cisgender woman (pronouns she/her) and an associate professor in SLP. She is the promotor (senior investigator) of this research and has 16 years of clinical and research experience in voice and 6 years in transgender voice. She is a member of a multidisciplinary gender team and the World and European Professional Association of Transgender Health. She was responsible for the conceptualization and methodology of the research project and was also part of the interpretation of the data and manuscript writing.

Participants of the RCT

Thirty-five transgender women were initially included in the study. Five participants dropped out in the course of the project, thus a total of 30 transgender women were included. They were recruited through the Gender team of the Ghent University Hospital (Belgium). The age of the participants ranged from 18 to 57 years, with a mean of 31 years (SD: 10.2). All participants had not yet received any speech training to feminize the voice. Inclusion criteria were an established diagnosis of gender dysphoria and female gender identity confirmed by the interdisciplinary Gender team at the Ghent University Hospital (Belgium), age between 18 and 70 years, self-reported normal hearing, Dutch speaker, with gender affirming hormonal treatment (both estrogens and anti-androgens, or after orchidectomy estrogens alone), a female gender role and seeking voice feminization care. Exclusion criteria were a history of neurological disorders, previous phonosurgery or voice and communication training to feminize the voice, organic pathology of the vocal folds, or smoking. Participants who smoked in the past but quit at least 1 month prior to the start of the training were not excluded.

Study design

This study used a randomized controlled trial with cross-over design. Participants were randomly assigned to a group and received 14 wk (= 15 h) of speech training. All participants started with 4 wk of sham training (1h15 per session), after which they were randomly assigned to one of two groups: one group continued with 5 wk of PET (1h per session), followed by 5 wk of ART (1h per session) and the second group received both trainings in the opposite order. Participants were recorded 4 times during the study, in between the training blocks: pre, post 1 (after sham), post 2 (after training 1) and post 3 (after training 2). This study aimed to 1. measure the impact of each program separately, i.e., sham training (pre—post 1), pitch elevation (pre—post 2) and articulation-resonance training (pre—post 2), and 2. measure the impact of the combination of all training programs (pre—post 3).

Speech intervention

All participants received the speech training in a sound treated room at Ghent University Hospital. The interventions were carried out by a certified and experienced speech language pathologist (CL). Sham training lasted for 4 wk (5 h), 1 session of 75 min per week, and included discussing vocal hygiene, anatomy, voice characteristics, non-verbal communication, relaxation and breathing exercises. Both the PET and ART lasted for 5 wk (5 h), 1 session of 60 min per week. A detailed description of the sessions of PET and ART can be found in Appendix A. Participants received a homework chart (Appendix B) where they could indicate whether they practiced or not. They were encouraged to exercise twice a day, 10 min each.

Data collection

Self-perception and satisfaction

The Trans Woman Voice Questionnaire (TWVQ) is a self-evaluation questionnaire that measures the transgender women's experiences with their voices (Dacakis et al., 2013). In this study, the Dutch (Flemish) version of the TVWQ was used (Van Borsel & Cosyns, 2016). The TWVQ consists of 30 statements about the current experience of one's own voice living as their authentic female gender. The 30 questions were evaluated according to a Likert scale (i.e., 1 = never or rarely, 2 = sometimes, 3 = often, 4 = usually or always).

The participants were also asked to rate their self-perception of the current and ideal voice concerning masculinity/femininity, changes in femininity after the sessions and satisfaction with the training program and results. These questions were asked with a VAS. The first two VAS consisted of rating the current and ideal masculinity-femininity of their voice, using the anchors "very masculine" (left side; score 0), "very feminine" (right side; score 100), and "neutral" in the middle. After interventions (i.e., all speech assessments except for the pre measurement), they were asked three more questions with a VAS, i.e., whether their voice sounded **more feminine** compared to their voice before starting therapy, whether they needed **more therapy** and whether they found the therapy **a pleasant and comfortable way to feminize their voice** ("strongly disagree" (left side), "strongly agree" (right side), "neutral" (middle)).

Both the TWVQ and VAS data were collected and managed using the REDCap (Research Electronic Data Capture) tools hosted at Ghent University. REDCap (Harris et al., 2009, 2019) is a secure, web-based software platform designed to support data capture for research studies.

Speech assessment

Speech samples. The RCT participants were recorded in a speech lab at Ghent University Hospital with a Samson C01U Pro USB Studio Condenser Microphone, digitized at a sampling rate of 44.1 kHz. The mouth-to-microphone distance was 15 cm during every recording. The samples contained continuous speech consisting of a phonetically balanced text "Papa en Marloes" (Van de Weijer & Slis, 1991). The first two sentences and the last sentence were cut from the text in order to keep a reasonable duration of the experiment. The remaining five sentences text were selected as speech sample for this experiment. Acoustic outcomes of this study were previously reported in Leyns et al. (2023).

Ten cisgender speakers (5 cisgender men, 5 cisgender women) were included in the listening experiment. Their speech samples, two per speaker, were collected to distract the listeners from the objective of the study. The ages of the cisgender speakers ranged from 24 to 66 years, with a mean age of 44 years (SD: 14.5). These speakers were recruited through convenience sampling, were all native Dutch speakers and had a self-reported normal hearing.

Listeners' perceptions

Listening experiments. Two listening experiments, continuous masculinity-femininity ratings and categorical masculinity-femininity attribution (resp. experiment A and B; explained below), were conducted using the online REDCap (Harris et al., 2009, 2019) tool. Listeners received the experiment as an online survey through this platform. They were instructed to complete the survey in a quiet room, using headphones. A test sample was included

to adjust the volume of the headphones to a comfortable intensity level. Twenty speech samples of cisgender men and women (2 samples per cisgender speaker) were also incorporated to distract the listeners from the objective of the study in order to avoid biased answers as much as possible. Ten double samples of the transgender women were added to calculate the intra-rater reliability. As the number of speech samples was too high to rate in one single moment due to a loss of concentration, the experiment was split into two rating rounds. Fifteen transgender women were rated during round one and fifteen during round two and all measurements of an individual were put in the same round. Listeners had to rate 150 speech samples in total (120 samples of transgender women (each participant providing 4 samples), 10 (2 times 5) samples for intra-rater reliability, 20 cisgender samples). Consequently, 75 samples were rated during the first round, and a second link was sent out seven days after the first one to complete the next 75 samples in round two. Half of the listeners started in the first round, half of them in the second round, and thus, received the next round in a different order. The anonymous samples were presented in random order and each listener started with a different speech sample. Even though the number of samples was not more than 75 samples per round, some people dropped out after week 1, resulting in missing data. However, half of the listeners started in the first round and half of them in the second round, resulting in an equal divide of missing data in both rounds.

Listening tasks. Experiment A investigated the masculinity-femininity on a VAS and experiment B on a categorical scale (M/F/X). In experiment A, listeners were instructed to rate the speech samples for masculinity-femininity on a VAS using the anchors "very masculine" (left side; score 0), "very feminine" (right side; score 100), and "neutral" in the middle. The listeners were instructed to treat the middle of the scale as ambiguous or neither feminine nor masculine. For experiment B, participants were asked to rate the person either "M," "F" ("V" in Dutch), or "X." On the general instruction page in the survey, these acronyms were described as "M = masculine, V = feminine, X = neutral/gender ambiguous." In order to distract the listeners from the objective and avoid biased answers, two extra questions (rating vocal quality and age on a VAS) were included in both experiments. The results of the two extra questions were not analyzed and were only used to prevent listeners from identifying the objectives of the study.

Listeners. In total, 75 listeners were included, 45 listeners participated in experiment A, and 30 listeners in experiment B. Cisgender, TGD and expert listeners were included. All participants had a self-reported normal hearing. They were all native Dutch speakers and were blinded to the purpose of the study and the gender(identity) of the participants.

Cisgender and TGD listeners ($n = 61$) were recruited as naïve listeners *via* convenience sampling. When asked about their knowledge of speech language pathology, they declared that they had no prior education or experience in this topic. Expert listeners ($n = 14$) were speech language pathologists who had experience with the topic of voice and transgender (having followed at least one specialist transgender voice course and guided at least 1 transgender client). Listener characteristics were displayed in Table 1.

Statistical analysis

SPSS 25.0 (SPSS Corp., Chicago, IL, USA) was used for the statistical analysis of the data. Analyses were conducted at $\alpha = 0.05$. A linear mixed model was used to compare the continuous outcome measures between the groups and between measurements at pre, post 1, post 2 and post 3. Time, Group and Time × Group interactions were specified as fixed factors. A random intercept for subjects was included and within-group effects of time were determined using pairwise comparisons with Bonferroni correction. Also, restricted maximum likelihood estimations and scaled identity covariance structures were used during the analyses. This covariance structure was chosen based on comparison of the Akaike's Information Criterion (AIC) values.

Inter-rater reliability was calculated for the listener experiments by means of two-way random ICCs, type consistency (single measures). The intra-rater reliability measured with two-way mixed ICCs, type absolute agreement (single

measures), was performed on ten double samples in the experiment for five randomly chosen raters. ICCs were interpreted following the classification of Altman (1990) (ICC < 0.20: poor, 0.21–0.40: fair, 0.41–0.60: moderate, 0.61–0.80: good, 0.81–1.00: very good).

Results

Self-perception and satisfaction

The evolution of the total scores of the TWVQ and VAS were reported in Tables 2 and 3. Sham training, comparing pre and post 1 measurements, did not show any significant differences. No significant differences have been observed in the TWVQ after each separate program. After the combination of both, the last measurement (post 3) showed a significantly lower value on the TWVQ total score, i.e., a less negative impact on the vocal functioning and the voice-related impact on their everyday life. On the VAS concerning the score of the current voice ("very masculine" (left side; score 0), "very feminine" (right side; score 100), and "neutral" in the middle), the participants scored their voices significantly more feminine after each separate program, and also after the combination of programs. No order effects were detected between ART and PET.

On Figure 1, the scores of their ideal and current voice were displayed. At the end of the therapy, this difference was smaller. Satisfaction questions had "strongly disagree" on the left side of the VAS (= 0), "strongly agree" on the right side (=100), and "neutral" in the middle (=50). The VAS with the question whether they think they actually sound more feminine after therapy, increased from 39 (SD 19.7, post 1) to 67 (SD 12.2, post 2) and 77 (SD 17.2, post 3). Participants agreed more at the end of the program that they do not need any more therapy, though, this mean score was still between "strongly disagree" and "neutral" (24 (SD 21.1), post 1; 32 (SD 25.1), post 2; 39 (SD 25.9), post 3). The scores on the question whether they found the program a fun and comfortable way to feminize the voice increased as well (70 (SD 17.0), post 1; 76 (SD 18.8), post 2; 86 (SD 14.1), post 3).

Table 1. Listeners in experiment A and B.

	Experiment A						Experiment B					
	Cisgender		TGD		Expert		Cisgender		TGD		Expert	
	PMAB	PFAB	PMAB	PFAB	PMAB	PFAB	PMAB	PFAB	PMAB	PFAB	PMAB	PFAB
N	16	11	5	3	1	9	10	8	6	2	0	4
Age (mean (SD, range))	45.1 (14.27, 26–69)	48.2 (13.27, 23–67)	25.8 (2.86, 22–29)	31.0 (13.86, 23–47)	29	37.3 (9.51, 26–51)	51.1 (15.85, 26–68)	45.0 (14.70, 26–66)	30. (11.49, 19–50)	49.0 (1.41, 48–50)		30.8 (5.91, 23–37)
Years of experience (mean (SD, range))					7	8.8 (6.30, 2–20)						2.2 (2.62, 0–6)
Total number of clients (mean (SD, range))					10	35.8 (38.92, 5–100)						5.3 (5.97, 1–14)

Abbreviations: PMAB, presumed male at birth; PFAB, presumed female at birth; SD, standard deviation.

Table 2. Self-perception and satisfaction of each separate program.

	PET				ART				PET vs. ART
	Pre (mean, SD)	Post 2 (mean, SD)	Mean difference	*p*-value	Pre (mean, SD)	Post 2 (mean, SD)	Mean difference	*p*-value	*p*-value
TWVQ total	68.5 (18.17)	62.8 (19.74)	5.7	0.312	79.2 (17.53)	74.5 (19.95)	4.7	0.613	0.805
Current voice (VAS)	22.7 (17.15)	44.6 (17.52)	21.9	**<0.001**	27.1 (16.34)	42.3 (16.49)	15.2	**0.001**	0.208
Ideal voice (VAS)	77.7 (16.16)	74.7 (13.16)	3.1	0.781	76.4 (8.72)	77.1 (8.40)	0.7	1.000	0.170

Bold values are significant (α = 0.05).

Table 3. Self-perception and satisfaction of combination of ART and PET.

	Pre (mean, SD)	Post 3 (mean, SD)	Mean difference	*p*-value	Order effect *p*-value
TWVQ total	73.9 (18.36)	65.9 (20.13)	8.0	0.001	0.115
Current voice (VAS)	24.9 (16.61)	49.5 (18.03)	24.6	<0.001	0.661
Ideal voice (VAS)	77.1 (12.78)	76.2 (14.01)	0.9	1.000	0.642

Abbreviations: PET, pitch elevation training; ART, articulation-resonance training; SD, standard deviation

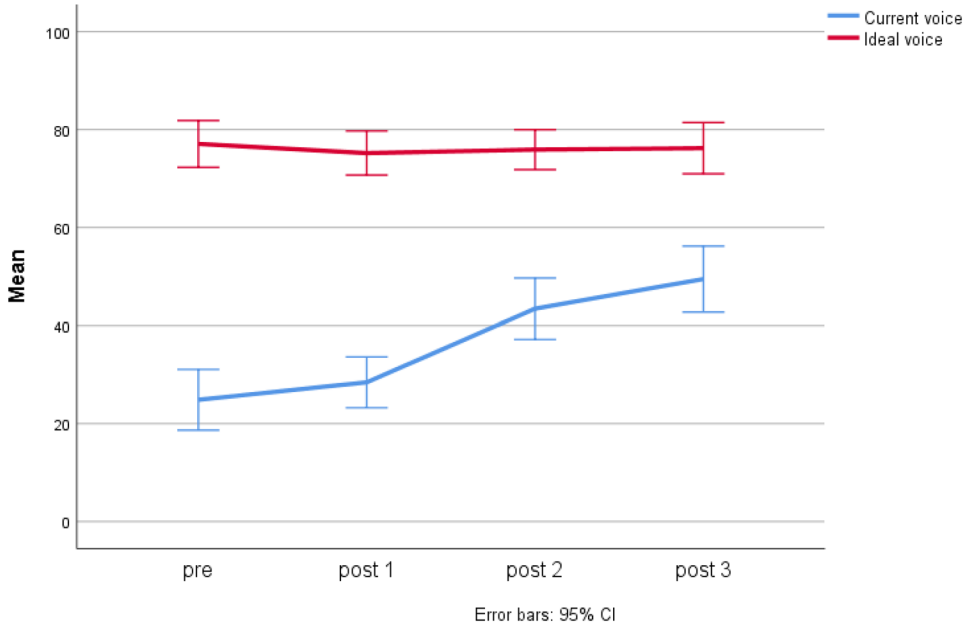


Figure 1. the current and ideal voice score on a visual analogue scale (0 = "very masculine," 100 = "very feminine").

Table 4. Listener perception of each separate program.

	PET				ART				PET vs. ART
	Pre (mean, SD)	Post 2 (mean, SD)	Mean difference	*p*-value	Pre (mean, SD)	Post 2 (mean, SD)	Mean difference	*p*-value	*p*-value
Masculinity—femininity (0–100mm)	27.3 (11.04)	45.1 (14.82)	17.9	**<0.001**	30.6 (14.72)	43.1 (16.28)	12.4	**<0.001**	0.175

Abbreviations: Pet, pitch elevation training; art, articulation-resonance training; sd, standard deviation.
Bold values are significant (α = 0.05).

Listening experiments

Experiment A

Listener reliability. The ICC for inter-rater reliability of experiment A was 0.72 (95% CI: 0.672–767). Concerning the intra-rater reliability, ICCs ranged between 0.65 and 0.89 for 5 raters (rater 1: 0.89 (95% CI: 0.628–0.973), rater 2: 0.87 (95% CI: 0.575–965), rater 3: 0.67 (95% CI: 0.070–906), rater 4: 0.65 (95% CI: 0.045–0.902), rater 5: 0.81 (95% CI: 0.402–0.949)). Following the classification of Altman (1990), inter-rater reliability was moderate, and intra-rater reliability ranges between moderate and very good.

Table 5. Listener perception of combination of ART and PET.

	Pre (mean, SD)	Post 3 (mean, SD)	Mean difference	*p*-value	Order effect *p*-value
Masculinity—femininity (0–100mm)	29.0 (12.90)	50.2 (14.71)	21.2	**<0.001**	0.822

Abbreviations: PET, pitch elevation training; ART, articulation-resonance training; SD, standard deviation.
Bold values are significant (α = 0.05).

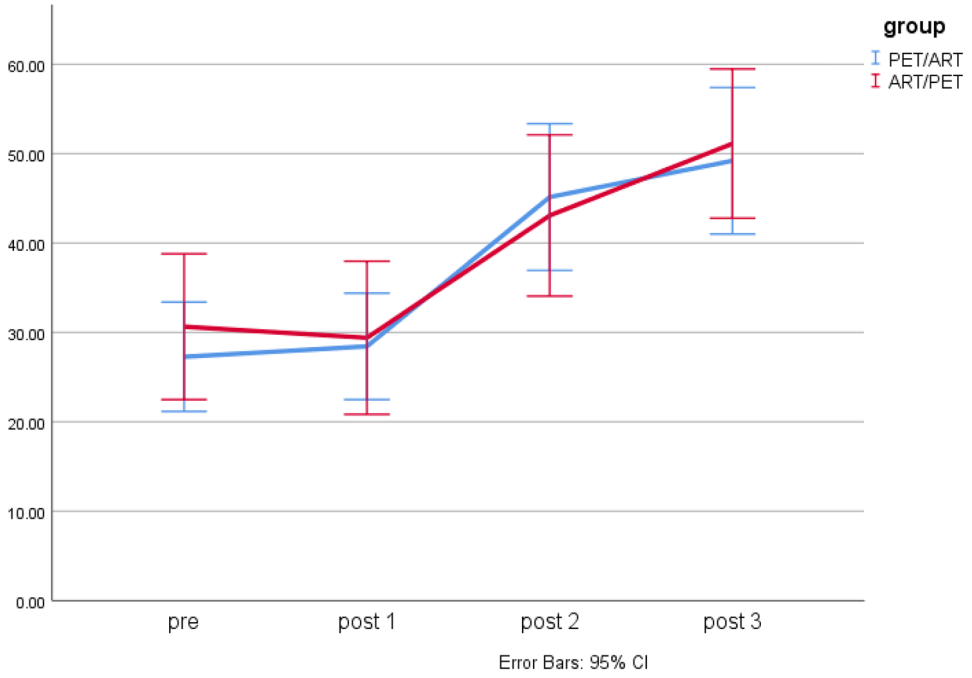


Figure 2. Group results during experiment A: masculinity/femininity score on a visual analogue scale (0 = "very masculine," 100 = "very feminine").

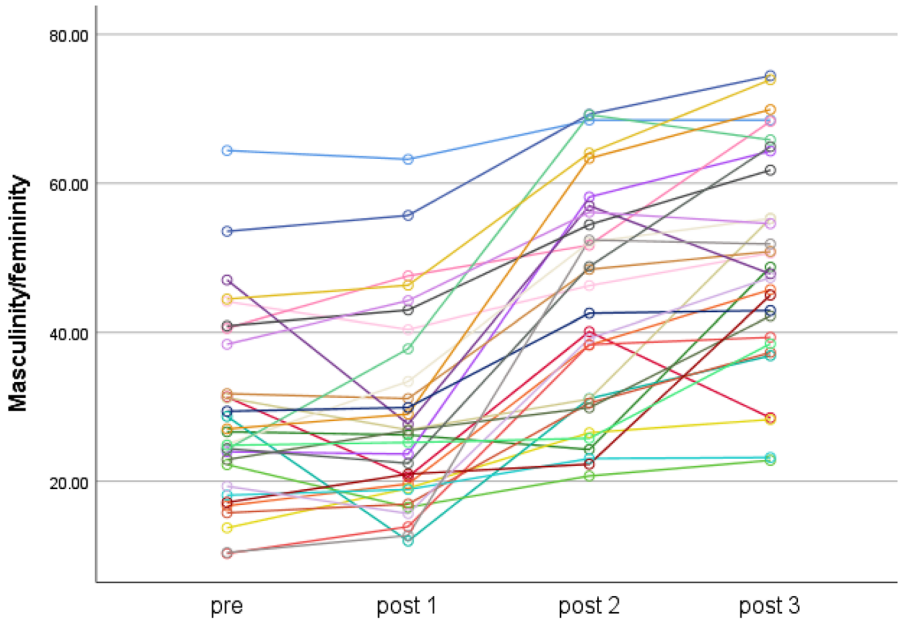


Figure 3. Individual results during experiment A: masculinity/femininity score on a visual analogue scale (0 = "very masculine," 100 = "very feminine").

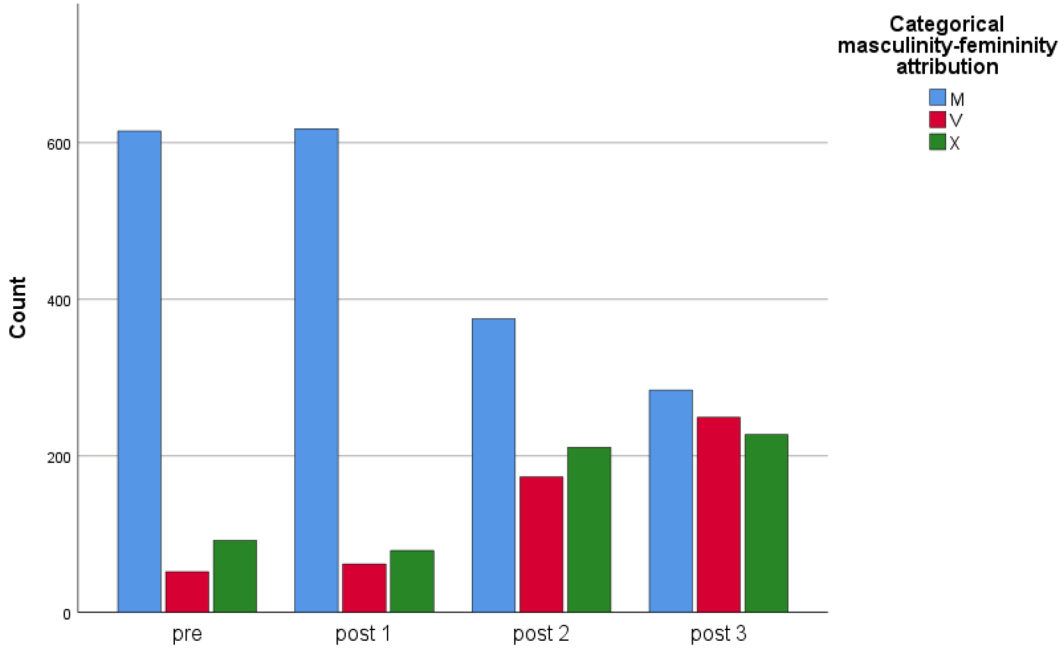


Figure 4. Group results during experiment B: categorical attribution (M/F/X, or M/V/X in Dutch).

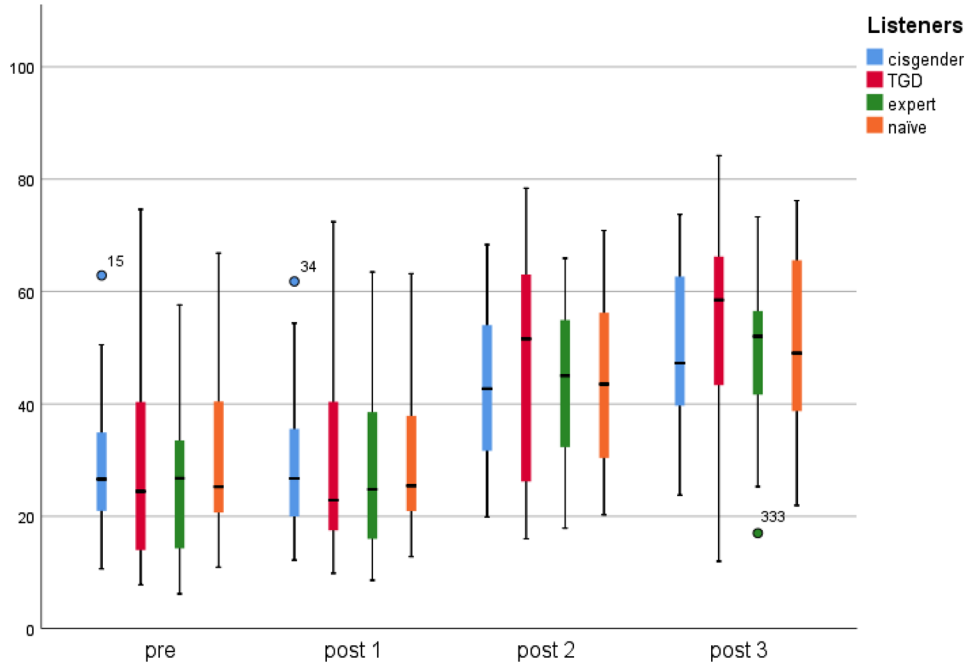


Figure 5. Listener group ratings over time (masculinity-femininity).

Continuous masculinity—femininity ratings. Participants' voices scored significantly closer to the feminine side of the VAS compared to the pre measurement, but still ranging around neutral, in the middle of the VAS. This was the case after each separate program (Table 4) and after the combination of programs (Table 5). Sham training, comparing pre and post 1 measurements, did not show any significant differences. No order effect was detected, thus, starting with PET or ART does not change the eventual outcome. Group and individual scores of all 30 participants were displayed on Figures 2 and 3.

Experiment B

Listener reliability. For experiment B, fair inter-rater reliability (0.41; 95% CI: 0.342–0.477), and

very good intra-rater reliability (1.00 for 4 raters, 0.94 for the fifth rater; 95% CI: 0.784–0.984) were found.

Categorical masculinity-femininity attribution. At the pre measurement, 81% of the participants were scored "M," 12.1% "X" and 6.9% "F." This remained stable at post 1, after the sham training (resp. 81.4%, 10.4% and 8.2%). After the first speech feminization training (either ART or PET), this changes to resp. 49.4%, 27.8% and 22.8%. At post 3, after the combination of programs, 37.4% were scored as "M," 29.9% as "X," and 32.8% as "F." These results were displayed on Figure 4.

Differences between listeners

When looking at the differences between the different listener groups (Figure 5), it is clear that the TGD group rated the samples with a more expansive range compared to the other groups. The median femininity score of the TGD group was higher at post 2 and 3 compared to the cisgender group. No differences were observed between experts and naïve listeners.

Discussion

The aim of this study was to measure and compare the short-term impact of both PET and ART in transgender women on self-perception, satisfaction and masculinity-femininity perceptions of listeners. Communicative dissatisfaction in transgender women due to inconsistent voice with the recognized gender is associated with symptoms such as depression and anxiety (Novais Valente Junior & Mesquita de Medeiros, 2022). How others perceive one's gender and femininity also is related to one's self-perceptions of femininity and happiness. McNeill et al. (2008) and Hancock et al. (2011) found that a listener's perception of femininity was positively correlated to the client's perception of her own femininity.

Concerning the self-perception during this study, the total score of the TWVQ showed that after the combination of both programs, a lower score, i.e., a lower negative impact on the vocal functioning and the voice-related impact on their everyday life, was detected. After ten sessions of speech feminization training, a mean difference of 8 points on the TWVQ was found between pre and post 3. It is not clear why the score did not change after the separate programs. The decreased score on the TWVQ after both could have been caused by the combination of all techniques, or the fact that more time has passed since the start of the training. In previous literature, data containing the self-perception and satisfaction of participants have been collected, either through an interview (Bralley et al., 1978; Quinn & Swain, 2018), VAS (Carew et al., 2007; Dacakis, 2000; Quinn & Swain, 2018; Söderpalm et al., 2004), Likert scales (Gelfer & Van Dong, 2013; Mészáros et al., 2005) and standardized questionnaires such as the Transsexual Self-Evaluation Questionnaire (TSEQ) by Davies and Goldberg (2006), (Hancock & Garabedian, 2013; Hancock & Helenius, 2012) and the Transsexual Voice Questionnaire (TVQ, more recently "Trans Woman Voice Questionnaire," TWVQ) by Dacakis et al. (2013), (Chadwick et al., 2022; Quinn & Swain, 2018). The studies reported increased satisfaction (Bralley et al., 1978; Carew et al., 2007; Dacakis, 2000; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Quinn & Swain, 2018), more self-confidence (Bralley et al., 1978; Hancock & Helenius, 2012), higher self-perception of femininity (Carew et al., 2007; Quinn & Swain, 2018; Söderpalm et al., 2004), lower scores on the TSEQ, TWVQ or Likert scale, i.e., lower negative impact on the psychosocial or voice-related functioning or limitations concerning communication (Chadwick et al., 2022; Hancock & Helenius, 2012; Mészáros et al., 2005; Quinn & Swain, 2018). In conclusion, it is hard to compare the current study with previous research, as different training programs, number of sessions, and overall methodology were used. When asking how their current voice and ideal voice (i.e., how feminine or masculine they want their voice to sound like), would sound like during the current study, changes have been detected as well. It is clear that over time, their perception of their ideal

voice stays stable as there were no significant time effects detected on the VAS about their ideal voice. Thus, they continued to have a similar goal in mind, but their current voice changed. A smaller difference between the two shows that they were in the process of achieving their goal of reaching a voice that is closer to their ideal voice, and thus, in better alignment with their gender identity. In contrast to the score on the TWVQ, this effect was observed after both separate programs, and the combination. Therefore, the participants felt that after 5 sessions their current voice was already closer to their ideal voice. As the study by Quinn and Swain (2018) mentioned, not all participants reached their goal of "a very feminine voice" (Quinn & Swain, 2018), but still made improvements in self-perception of femininity of the voice from some negative feelings they experienced before the training. Intervention that targets the client's self-perception of the voice in addition to focusing on the acoustic aspects such as f_o during speaking and resonance is recommended for the development of a feminine voice (Hancock & Helenius, 2012). Vocal satisfaction and self-perception of femininity of the voice might be a feasible treatment goal in itself, independent of a client's goals around passing, i.e., being perceived as their true gender in everyday life, and socializing (Quinn & Swain, 2018).

The other VAS showed that they found their voices sounding more feminine after therapy and that they experienced the sessions as a comfortable and fun way to feminize their voices. Concerning the question if they do not need more therapy, scores were more closely to "strongly disagree" and "neutral," demonstrating that some might need more support from an SLP after the ten sessions. It is possible that they were not completely comfortable or still insecure with using their feminine voice, or that they wanted to generalize more techniques to daily speech. Additionally, it takes time to adjust their idea about the "new voice" and to get used to their voice themselves based on new communication experiences. However, this was not examined in this study. A qualitative study design is necessary in the future to gain awareness about these aspects.

In order to gain insights into perceptions of listeners, two different listening experiments were conducted: one listening panel scored on a VAS, ranging between very masculine to very feminine with neutral in the middle. The second panel scored the categorical masculinity-femininity attribution in a trichotomous way, i.e., "M" (masculine), "F" (feminine), and "X" (neutral or gender ambiguous). As all of the transgender participants had a goal of feminizing the voice, including a categorical attribution experiment is useful as it might reflect to a certain degree how they are perceived in daily life, and whether occasions of misgendering occur. Although the authors are aware that voice (identity) is a spectrum and is difficult to divide into a categorical variable (M/F/X), it is still important for some transgender clients to be perceived as feminine and not as ambiguous.

Listener reliabilities were measured for both experiments and were found to be moderate (inter) and moderate to very good (intra) during experiment A, and fair (inter) and very good (intra) during experiment B. Low inter-rater reliabilities were to be expected based on the study by Azul and Hancock (2020), based on the different constructs that listeners have about attributing gender to a certain voice. Even if an ICC is interpreted as "very good," one should consider that some listeners might not have felt confident in their judgments. In a study by Honorof and Whalen (2010) it was found that listeners were less confident when rating audio recordings that where in a "midrange" (i.e., gender ambiguous) f_o zone. As most transgender participants were in that zone (Leyns et al., 2023), the listeners' lack of confidence while answering the questions might have biased the results; however this was not analyzed in this study. Furthermore, it is harder to achieve a good inter-rater reliability during categorical ratings compared to continuous ratings. Carew et al. (2007) reported poor inter-rater reliability during their listening experiment and therefore suggested to use speech pathologists to carry out perceptual ratings. The

speech samples in our current study were evaluated by a listening panel consisting of cisgender, transgender and non-binary listeners and also including both expert listeners (SLPs) and naïve listeners. In previous studies listening panels typically consisted of cisgender persons. One might think that TGD listeners would evaluate the speech samples differently based on the idea that they are more aware about their own speech than cisgender persons (Brown et al., 2021). Although Brown et al. (2021) and Quinn et al. (2022) found no influence of rater's gender on ratings of gender perception by listeners, it is still important to be aware for diversity within a panel of listeners. In the current study, the range of the TGD panel is larger compared to the cisgender group (Figure 5), and their median score was also higher. This corresponds with what has been found in the study by Hope and Lilley (2022), where ranges were greatest in the TGD group. The authors mentioned that this would indicate a greater distribution as they use the full scale and not only the extremes. In their study, Hope and Lilley (2022) confirmed that their findings were parallel with previous research stating that the characteristics of the perceiver contribute more to forming impressions than the appearance or presentation of the target (Xie et al., 2019), and that a person's conceptual knowledge of self is reflected in their social judgements of others (Stolier et al., 2020).

When looking at the masculinity—femininity ratings, transgender women's voices were scored 18% more feminine after PET, and 12% after ART. However, it is important to acknowledge that the mean group values (resp. pre 27 mm—post 2 45 mm, pre 31 mm—post 2 43 mm) were still in the left side of the VAS, thus ranging between masculine and neutral. Concerning the combination of both programs, a mean difference of 21% was found (pre 29 mm—post 3 50 mm) for the combined training. Starting with PET or ART did not change the outcome at post 3. Individual results concerning masculinity/femininity were also displayed. In general, the score is stable after the sham training. Many participants make a steep increase toward post 2, after the first speech feminization intervention. All participants ended up with a higher femininity perception at post 3 compared to the pre measurement. This finding alludes to the fact that for some people, specific speech feminization techniques work better for some people compared to others and that an individual approach is necessary. In previous studies, transgender women's voices were reported more feminine after training on a VAS (Bralley et al., 1978; Carew et al., 2007; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock & Helenius, 2012; Kaye et al., 1993; Quinn & Swain, 2018). Similar to these studies, the current study also investigated the effects of a combined therapy program. Unlike the study by Leyns, Corthals, et al. (2021), who researched the effect of a single exercise, i.e., lip spreading or the cork exercise. Therefore, it is difficult to draw conclusions on which training technique induced a certain effect on self-perception, satisfaction, or listener perceptions.

During the categorical masculinity-femininity attribution experiment, listeners were asked to rate each speech sample either "M," "F" or "X." A shift toward neutral/gender ambiguous and feminine categories has been observed. Sham therapy did not result in changes of attribution. At post 3, still 37% of the individuals were rated as masculine, and only 30% neutral and 33% feminine. Coming from 81% masculine at the pre measurement, one might say that some progress has been made, although more than one third of the individuals might still be misgendered by listeners. Ten sessions of 60 min once a week were included in the current study. One should ask why the percentage of feminine attributions is still quite low after this therapy program. It is unclear whether the content, duration or frequency of the program, home exercises, self-efficacy, or other background factors might contribute to these results. Defining outcome predictors in order to examine how feminine a transgender woman will be perceived after training is crucial in future research. Recently, Dacakis et al. (2022) developed a self-efficacy scale for voice modification related to gender affirmation. In previous research, it was demonstrated that self-efficacy was a "clinically meaningful" predictor of adherence to therapy and in

particular, home practice (van Leer & Connor, 2012). Such a scale could be used to inform voice training goals and explore predictors of voice training outcomes.

As previously mentioned, SLPs should include counseling about the unpredictability of voice production and speaker socio-cultural positioning in conversations (Azul & Hancock, 2020). Also, the SLP can help with developing confident responses to misattributions from conversation partners such as misgendering (Azul & Hancock, 2020). Although they are difficult to compare as they used different methodologies, experiments with binary attribution ratings in previous research revealed a higher number of female attributions post treatment and follow-up in the study by Gelfer and Tice (2013). In the study by Gelfer and Van Dong (2013) and Kaye et al. (1993), many participants were still rated as male following training. It might be possible that other vocal factors need to change in order to achieve more feminine voice ratings during the current and past studies. Acoustic outcomes (Leyns et al., 2023) should be compared with listener perceptions, self-perception and satisfaction. It is important to consider additional analysis of listeners' backgrounds, because listeners can be differently socialized and have different constructions of gender or masculinity-femininity attributions. A person's experience with the category of gender and their training or habit related to attributing gender to voices should bring insights to the results of listening experiments. As this information was not collected during the current study, this is a clear limitation during the interpretation of the data. In fact, it is also hard to investigate gender in isolation from other aspects of socio-cultural positioning in this study. In order to take into account a person's socio-cultural positioning, qualitative research can provide deeper insights into the personal experiences and perceptions of these individuals. They will help the SLP to better understand the types of forces that act on voice production and of these forces' capacity to influence meaning-making based on voice (Azul & Hancock, 2020). Researcher positionings have been added in this study, which showed that during the interpretation of data, no TGD researchers were included. This is also an apparent limitation and should definitely be accounted for in future research.

Conclusion

This study measured and compared the short-term impact of both PET and ART in transgender women on self-perception, satisfaction and masculinity-femininity perceptions of listeners. Concerning the self-perception of the transgender women, higher femininity was detected after the sessions. The TWVQ showed a lower negative voice-related impact on their daily life and vocal functioning. When looking at the perceptions of listeners, participants' voices were perceived more feminine, both detected on a continuous visual analogue scale and during categorical attribution. The combination of both programs (ART and PET) revealed higher femininity for self-perception and perception by listeners. However, a lot of the participants were still perceived masculine at the end of the program. Defining outcome predictors from a client's background or certain vocal characteristics in order to examine how feminine a transgender woman will be perceived after training is crucial in future research.

Disclosure statement

The authors declare that they have no conflict of interest.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, *35*(4), 492–503. https://doi.org/10.1080/03637756809375599

Altman, D. G. (1990). *Practical statistics for medical research* (1st ed.). Chapman & Hall.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, *70*(9), 832–864. https://doi.org/10.1037/a0039906

Azul, D., & Hancock, A. B. (2020). Who or what has the capacity to influence voice production? Development of a

transdisciplinary theoretical approach to clinical practice addressing voice and the communication of speaker socio-cultural positioning. *International Journal of Speech-Language Pathology*, *22*(5), 559–570. https://doi.org/10.1080/17549507.2019.1709544

Azul, D., Hancock, A. B., Lundberg, T., Nygren, U., & Dhejne, C. (2022). Supporting well-being in gender-diverse people: A tutorial for implementing conceptual and practical shifts toward culturally responsive, person-centered care in speech-language pathology. *American Journal of Speech-Language Pathology*, *31*(4), 1574–1587. https://doi.org/10.1044/2022_ajslp-21-00322

Bralley, R. C., Bull, G. L., Gore, C. H., & Edgerton, M. T. (1978). Evaluation of vocal pitch in male transsexuals. *Journal of Communication Disorders*, *11*(5), 443–449. https://doi.org/10.1016/0021-9924(78)90037-0

Brown, K. M., Dahl, K. L., Cler, G. J., & Stepp, C. E. (2021). Listener age and gender diversity: Effects on voice-based perception of gender. *Journal of Voice*, *35*(5), 739–745. https://doi.org/10.1016/j.jvoice.2020.02.004

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice*, *21*(5), 591–603. https://doi.org/10.1016/j.jvoice.2006.05.005

Cartei, V., Bond, R., & Reby, D. (2014). What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men's vocal masculinity. *Hormones and Behavior*, *66*(4), 569–576. https://doi.org/10.1016/j.yhbeh.2014.08.006

Chadwick, K. A., Coleman, R., Andreadis, K., Pitti, M., & Rameau, A. (2022). Outcomes of gender-affirming voice and communication modification for transgender individuals. *The Laryngoscope*, *132*(8), 1615–1621. https://doi.org/10.1002/lary.29946

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., Fraser, L., Green, J., Knudson, G., Meyer, W. J., Monstrey, S., Adler, R. K., Brown, G. R., Devor, A. H., Ehrbar, R., Ettner, R., Eyler, E., Garofalo, R., Karasic, D. H., … Zucker, K. (2012). World Professional Association for Transgender Health. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *International Journal of Transgenderism*, *13*(4), 165–232. https://doi.org/10.1080/15532739.2011.700873

Coleman, E., Radix, A. E., Bouman, W. P., Brown, G. R., de Vries, A. L. C., Deutsch, M. B., Ettner, R., Fraser, L., Goodman, M., Green, J., Hancock, A. B., Johnson, T. W., Karasic, D. H., Knudson, G. A., Leibowitz, S. F., Meyer-Bahlburg, H. F. L., Monstrey, S. J., Motmans, J., Nahata, L., … Arcelus, J. (2022). Standards of care for the health of transgender and gender diverse people, version 8. *International Journal of Transgender Health*, *23*(Suppl 1), S1–S259. https://doi.org/10.1080/26895269.2022.2100644

Dacakis, G. (2000). Long-term maintenance of fundamental frequency increases in male-to-female transsexuals. *Journal of Voice*, *14*(4), 549–556. https://doi.org/10.1016/s0892-1997(00)80010-7

Dacakis, G., Davies, S., Oates, J., Douglas, J., & Johnston, J. R. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *Journal of Voice*, *27*(3), 312–320. https://doi.org/10.1016/j.jvoice.2012.11.005

Dacakis, G., Erasmus, J., Nygren, U., Oates, J., Quinn, S., & Södersten, M. (2022). Development and initial psychometric evaluation of the self-efficacy scale for voice modification in trans women. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2022.03.015

Davies, S., & Goldberg, J. M. (2006). Clinical aspects of transgender speech feminization and masculinization. *International Journal of Transgenderism*, *9*(3–4), 167–196. https://doi.org/10.1300/J485v09n03_08

Davies, S., Papp, V. G., & Antoni, C. (2015). Voice and communication change for gender nonconforming individuals: Giving voice to the person inside. *International Journal of Transgenderism*, *16*(3), 117–159. https://doi.org/10.1080/15532739.2015.1075931

Gelfer, M. P., & Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to--female transsexuals perceived as female versus those perceived as male. *Journal of Voice*, *14*(1), 22–33. https://doi.org/10.1016/S0892-1997(00)80092-2

Gelfer, M. P., & Tice, R. M. (2013). Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later. *Journal of Voice*, *27*(3), 335–347. https://doi.org/10.1016/j.jvoice.2012.07.009

Gelfer, M. P., & Van Dong, B. R. (2013). A preliminary study on the use of vocal function exercises to improve voice in male-to-female transgender clients. *Journal of Voice*, *27*(3), 321–334. https://doi.org/10.1016/j.jvoice.2012.07.008

Grant, J. M., Motter, L. A., & Tanis, J. (2011). *Injustice at every turn: A report of the national transgender discrimination survey*. National Center for Transgender Equality and National Gay and Lesbian Task Force.

Haas, A. P., Rodgers, P. L., & Herman, J. L. (2014). *Suicide attempts among transgender and gender non-conforming adults* (Vol. 50). Amercian Foundation for Suicide Prevention.

Hancock, A. B., & Garabedian, L. M. (2013). Transgender voice and communication treatment: A retrospective chart review of 25 cases. *International Journal of Language & Communication Disorders*, *48*(1), 54–65. https://doi.org/10.1111/j.1460-6984.2012.00185.x

Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal of Voice*, *25*(5), 553–558. https://doi.org/10.1016/j.jvoice.2010.07.013

Hancock, A., & Helenius, L. (2012). Adolescent male-to-female transgender voice and communication therapy. *Journal of Communication Disorders*, *45*(5), 313–324. https://doi.org/10.1016/j.jcomdis.2012.06.008

Hancock, A., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice*, *28*(2), 203–209. https://doi.org/10.1016/j.jvoice.2013.08.009

Hardy, T. L. D., Rieger, J. M., Wells, K., & Boliek, C. A. (2020). Acoustic predictors of gender attribution, masculinity-femininity, and vocal naturalness ratings amongst transgender and cisgender speakers. *Journal of Voice*, *34*(2), 300.e311–300.e326. https://doi.org/10.1016/j.jvoice.2018.10.002

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Honorof, D. N., & Whalen, D. (2010). Identification of speaker sex from one vowel across a range of fundamental frequencies. *The Journal of the Acoustical Society of America*, *128*(5), 3095–3104. https://doi.org/10.1121/1.3488347

Hope, M., & Lilley, J. (2022). Gender expansive listeners utilize a non-binary, multidimensional conception of gender to inform voice gender perception. *Brain and Language*, *224*, 105049. https://doi.org/10.1016/j.bandl.2021.105049

Kanamori, Y., & Xu, Y. J. (2022). Factors associated with transphobia: A structural equation modeling approach. *Journal of Homosexuality*, *69*(4), 716–740. https://doi.org/10.1080/00918369.2020.1851959

Kaye, J., Bortz, M. A., & Tuomi, S. K. (1993). Evaluation of the effectiveness of voice therapy with a male-to-female transsexual subject. *Scandinavian Journal of Logopedics and Phoniatrics*, *18*(2-3), 105–109. https://doi.org/10.3109/14015439309101356

Klemmer, C. L., Arayasirikul, S., & Raymond, H. F. (2021). Transphobia-based violence, depression, and anxiety in transgender women: The role of body satisfaction. *Journal of Interpersonal Violence*, *36*(5-6), 2633–2655. https://doi.org/10.1177/0886260518760015

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, *59*(3), 675–678. https://doi.org/10.1121/1.380917

Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research : JSLHR*, *61*(2), 266–297. https://doi.org/10.1044/2017_jslhr-s-17-0067

Leung, Y., Oates, J., Chan, S. P., & Papp, V. (2021). Associations between speaking fundamental frequency, vowel formant frequencies, and listener perceptions of speaker gender and vocal femininity–masculinity. *Journal of Speech, Language, and Hearing Research : JSLHR*, *64*(7), 2600–2622. https://doi.org/10.1044/2021_JSLHR-20-00747

Leyns, C., Corthals, P., Cosyns, M., Papeleu, T., Van Borsel, J., Morsomme, D., T'Sjoen, G., & D'haeseleer, E. (2021). Acoustic and perceptual effects of articulation exercises in transgender women. *Journal of Voice*, https://doi.org/10.1016/j.jvoice.2021.06.033

Leyns, C., Daelman, J., Adriaansen, A., Tomassen, P., Morsomme, D., T'Sjoen, G., & D'haeseleer, E. (2023). Short-term acoustic effects of speech therapy in transgender women: A randomized controlled trial. *American Journal of Speech-Language Pathology*, *32*(1), 145–168. https://doi.org/10.1044/2022_AJSLP-22-00135

Leyns, C., Papeleu, T., Tomassen, P., T'Sjoen, G., & D'haeseleer, E. (2021). Effects of speech therapy for transgender women: A systematic review. *International Journal of Transgender Health*, *22*(4), 1–21. https://doi.org/10.1080/26895269.2021.1915224

McNeill, E. J., Wilson, J. A., Clark, S., & Deakin, J. (2008). Perception of voice in the transgender client. *Journal of Voice*, *22*(6), 727–733. https://doi.org/10.1016/j.jvoice.2006.12.010

Mészáros, K., Vitéz, L. C., Szabolcs, I., Góth, M., Kovács, L., Görömbei, Z., & Hacki, T. (2005). Efficacy of conservative voice treatment in male-to-female transsexuals. *Folia Phoniatrica Et Logopaedica*, *57*(2), 111–118. https://doi.org/10.1159/000083572

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin*, *129*(5), 674–697. https://doi.org/10.1037/0033-2909.129.5.674

Millet, N., Longworth, J., & Arcelus, J. (2017). Prevalence of anxiety symptoms and disorders in the transgender population: A systematic review of the literature. *International Journal of Transgenderism*, *18*(1), 27–38. https://doi.org/10.1080/15532739.2016.1258353

Novais Valente Junior, C., & Mesquita de Medeiros, A. (2022). Voice and gender incongruence: Relationship between vocal self-perception and mental health of trans women. *Journal of Voice*, *36*(6), 808–813. https://doi.org/10.1016/j.jvoice.2020.10.002

Oates, J. (2019). Evidence-based practice in voice training for trans women. In R. K. Adler, S. Hirsch, & J. Pickering (Eds.), *Voice and communication therapy for the transgender/gender diverse client: A comprehensive clinical guide* (3rd ed., pp. 87–103). Plural Publishing Inc.

Quinn, S., & Swain, N. (2018). Efficacy of intensive voice feminisation therapy in a transgender young offender. *Journal of Communication Disorders*, *72*, 1–15. https://doi.org/10.1016/j.jcomdis.2018.02.001

Quinn, S., Oates, J., & Dacakis, G. (2022). Perceived gender and client satisfaction in transgender voice work: Comparing self and listener rating scales across a training program. *Folia Phoniatrica Et Logopaedica*, *74*(5), 364–379. https://doi.org/10.1159/000521226

Schneider, M. A., Spritzer, P. M., Soll, B. M. B., Fontanari, A. M. V., Carneiro, M., Tovar-Moll, F., Costa, A. B., da

Silva, D. C., Schwarz, K., Anes, M., Tramontina, S., & Lobato, M. I. R. (2017). Brain maturation, cognition and voice pattern in a gender dysphoria case under pubertal suppression. *Frontiers in Human Neuroscience*, *11*, 528. https://doi.org/10.3389/fnhum.2017.00528

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Söderpalm, E., Larsson, A., & Almquist, S.-Å. (2004). Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logopedics, Phoniatrics, Vocology*, *29*(1), 18–30. https://doi.org/10.1080/14015430310021618

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371. https://doi.org/10.1038/s41562-019-0800-6

T'Sjoen, G., Arcelus, J., Gooren, L., Klink, D. T., & Tangpricha, V. (2019). Endocrinology of Transgender Medicine. *Endocrine Reviews*, *40*(1), 97–117. https://doi.org/10.1210/er.2018-00011

Tan, K. K., Treharne, G. J., Ellis, S. J., Schmidt, J. M., & Veale, J. F. (2020). Gender minority stress: A critical review. *Journal of Homosexuality*, *67*(10), 1471–1489. https://doi.org/10.1080/00918369.2019.1591789

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), 65–77. https://doi.org/10.1037/sgd0000081

Van Borsel, J., & Cosyns, M. (2016). Dutch (Flemish) version of the trans woman voice questionnaire. In *Vakgroep Spraak-, Taal- en Gehoorwetenschappen*. Ghent University.

Van de Weijer, J., & Slis, I. (1991). Nasaliteitsmeting met de nasometer. *Logopedie En Foniatrie*, *63*, 97–101.

van Leer, E., & Connor, N. P. (2012). Use of portable digital media players increases patient motivation and practice in voice therapy. *Journal of Voice*, *26*(4), 447–453. https://doi.org/10.1016/j.jvoice.2011.05.006

Villas-Bôas, A. P., Schwarz, K., Fontanari, A. M. V., Costa, A. B., Cardoso da Silva, D., Schneider, M. A., Cielo, C. A., Spritzer, P. M., & Rodrigues Lobato, M. I. (2021). Acoustic measures of brazilian transgender women's voices: A case-control study. *Frontiers in Psychology*, *12*, 622526. https://doi.org/10.3389/fpsyg.2021.622526

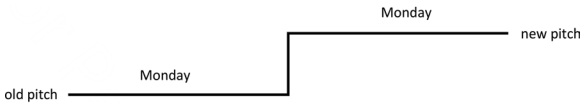
Vincent, B. (2018). *Transgender health: A practitioner's guide to binary and non-binary trans patient care*. Jessica Kingsley Publishers.

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. https://doi.org/10.1037/pspi0000160

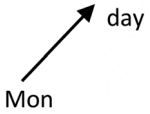
Appendix A

1. Pitch elevation training
   a. Session 1
      i. Auditory descrimination with a piano
      ii. Glissando patterns (using biofeedback real-time pitch of Computerized Speech Lab (Kay Elemetrics)
         1. From habitual "old" pitch to the "new," higher pitch in isolated nasal consonants (approximately till 160 Hz); from habitual old pitch to highest pitch, etc.
      iii. Adding consonant-vowel-consonant combinations
      iv. Explanation of biofeedback tool to use at home: smartphone app Voice Pitch Analyzer

   b. Session 2
      i. Repetition of glissando patterns
      ii. Automatic sequences (counting from 1 till 10, days of the week, months of the year, etc.) starting on habitual old pitch and then on the new pitch



      iii. Automatic sequences with gliding from old to new pitch within the word



   c. Session 3
      i. Short warm-up with glissando patterns
      ii. Speaking with the new pitch, making sure there are a lot of upward intonation patterns
         1. Short expressions (e.g., "Go away," "Mum and dad," "up and down," "be careful," etc.)
         2. Building up sentences
            Can you
            Can you put
            Can you put those files
            Can you put those files in the storage room later on?
         2. Short sentences
         3. Poems
         4. Texts

   d. Session 4
      i. Introduction of water resistance therapy, using a resonance tube (2 cm under water)
         1. Bubbeling without phonation
         2. Phonation with old pitch in the tube
         3. Phonation with new pitch in the tube
         4. Glissando patterns in the tube
         5. Short sentences with new pitch in the tube, then without tube
         6. Poems with new pitch in the tube, then without tube

7. Texts with new pitch in the tube, then without tube
8. Spontaneous speech, answering in the tube, then without tube

a. Short answers (1 sentence)
b. Longer answers (2–3 sentences)
c. Conversation

e. Session 5
   i. Repetition of water resistance therapy
   ii. Straw phonation
      1. Blowing without phonation
      2. Phonation on new pitch
      3. Glissando patterns
      4. Spontaneous speech, answering in the straw, then without straw

2. Articulation-resonance training
   a. Session 1
      i. Lip spreading
         1. Alternating with making an /u/ and /i/ movement of the lips (= discrimination with lip protrusion and lip spreading). Using a mirror to look at the lip movements.
         2. Alternating with making an /u/ and /i/ sound.
         3. Alternating with making an /e/ and /y/ sound.
         4. Consonant + /i/ combinations, consonant + /e/ combinations, feeling the easy lip spreading
         5. Trying to reach lip spreading when doing consonant + /u/ and /y/ combinations
         6. Monosyllable words with /i/, /e/, /u/ and /y/
         7. Multisyllable words with /i/, /e/, /u/ and /y/
         8. Sentences with /i/
         9. Sentences with /e/
         10. Sentences with all combinations
         11. Text
         12. Spontaneous speech

   b. Session 2
      i. Repetition of lip spreading
      ii. Forward tongue position
         1. Awareness of the tongue muscle: non-speech oral motor exercises
            Using a mirror to look at the tongue movements
         2. Moving the tongue from front to back when producing vowels
         3. Pronouncing /i/ (high vowel) and feeling the forward tongue position with a high back of the tongue
         4. Starting from /i/ sound and gliding to other vowels, trying to reach forward tongue and high back of the tongue
         5. Words with /i/ (high vowel)
         6. Words with /y/ (high vowel)
         7. Words with /a/ (low vowel)
         8. Sentences with /a/
         9. Words with /o/ (low vowel)
         10. Sentences with /o/
         11. Texts
         12. Spontaneous speech

   c. Session 3
      i. Repetition of forward tongue position
      ii. Larynx elevation through twang
         1. Awareness exercise: yawning (downward movement of the larynx) and swallowing (upward movement of the larynx)
         2. Listening to twang sound such as crying baby, goat sounds, etc.
         3. Adding twang to vowel /a/
         4. Decreasing twang to vowel /a/
         5. Consonant + /a/ + consonant + /a/ + consonant + /a/ with twang
         6. Words with /a/
         7. Sentences with /a/
         8. Texts
         9. Spontaneous speech

   d. Session 4
      i. Repetition of larynx elevation
      ii. Forward resonance
         1. Discrimination between chest resonance and head resonance, saying /o/ vowel
         2. Putting a finger on left and right nostril and saying "hmmm," feeling forward airflow
         3. Nasal consonant /m/ + vowel
         4. Words with initial /m/
         5. Extra exercise to feel forward resonance
            a. Stand in front of a wall about 50 cm away.
            b. Place your head against the wall, comfortably.
            c. Your arms hang loose by your side.
            d. Place your tongue on your hard palate and start with a "nnnn" sound. Make a few glissandos to high and low frequencies.
            e. Repeat the previous step but now place the back part of your tongue on your soft palate. Make a "ng" sound and a few glissandos. By placing the head against the wall you feel the resonances better in your head.

      iii. Clear speech
         1. Combinations of consonants and vowels, pronouncing slow and then very fast, trying to pronounce clearly and precisely
            a. Tippetiptiptip tappetaptaptap toppetoptoptop
            b. Tanatanta tenetente tinitinti tonotonto
            c. Prieke prokke prakke pro prieke prokke prakke pro
            d. …
         2. Word combinations, 3x slow and 3x fast
         3. Cork exercise: using a cork with a diameter of 23 mm and length of 45 mm.

a. Placing the upper front of the cork (approximately 2–3 mm) between their front teeth and reading words out loud with large and precise articulation movements. After a block of long nouns (6–9 syllables), they removed the cork and used the same large articulation movements to pronounce the same block of words.
b. Tongue twisters with and without cork
c. Text: reading sentences with and without cork

4. Spontaneous speech

e. Session 5
i. Repetition of all articulation-resonance techniques, spending most time on forward resonance and clear speech
ii. Generalization of all articulation-resonance techniques
1. Texts
2. Spontaneous speech

Appendix B

Oefenschema (Homework chart)
Teken een smiley bij de dagen wanneer je geoefend hebt. (draw a smiley/circle/cross on the days that you practiced)
Ideaal: 2x per dag telkens 10 min. (ideal: 2x a day for 10 min each)

	Maandag (Monday)	Dinsdag (Tuesday)	Woensdag (Wednesday)	Donderdag (Thursday)	Vrijdag (Friday)	Zaterdag (Saturday)	Zondag (Sunday)
Week 5							
Week 6							
Week 7							
Week 8							
Week 9							
	Maandag *(Monday)*	Dinsdag *(Tuesday)*	Woensdag *(Wednesday)*	Donderdag *(Thursday)*	Vrijdag *(Friday)*	Zaterdag *(Saturday)*	Zondag *(Sunday)*
Week 10							
Week 11							
Week 12							
Week 13							
Week 14							