International Journal of Transgender Health

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/wijt21

Examining the voice of Israeli transgender women: Acoustic measures, voice femininity and voice-related quality-of-life

Noa Diamant & Ofer Amir

To cite this article: Noa Diamant & Ofer Amir (2020): Examining the voice of Israeli transgender women: Acoustic measures, voice femininity and voice-related quality-of-life, International Journal of Transgender Health, DOI: 10.1080/26895269.2020.1798838

To link to this article: https://doi.org/10.1080/26895269.2020.1798838

Published online: 07 Aug 2020.

INTERNATIONAL JOURNAL OF TRANSGENDERISM
https://doi.org/10.1080/26895269.2020.1798838

Examining the voice of Israeli transgender women: Acoustic measures, voice femininity and voice-related quality-of-life

Noa Diamant and Ofer Amir

Department of Communication Disorders, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

ABSTRACT

Background: Transgender women may experience gender-dysphoria associated with their voice and the way it is perceived. Previous studies have shown that specific acoustic measures are associated with the perception of voice-femininity and with voice-related quality-of-life, yet results are inconsistent.

Aims: This study aimed to examine the associations between specific voice measures of transgender women, voice-related quality-of-life, and the perception of voice-femininity by listeners and by the speakers themselves.

Methods: Thirty Hebrew speaking transgender women were recorded. They had also rated their voice-femininity and completed the Hebrew version of the TVQ^{MtF} questionnaire. Recordings were analyzed to extract mean fundamental frequency (F0), formant frequencies (F1, F2, F3), and vocal-range (calculated in Hz. and in semitones). Recordings were also rated on a voice-gender 7-point scale, by 20 naïve cisgender listeners.

Results: Significant correlations were found between both F0 and F1 and listeners' as well as speakers' evaluation of voice-femininity. TVQ^{MtF} scores were significantly correlated with F0 and with the lower and upper boundaries of the vocal-range. Voice-femininity ratings were strongly correlated with vocal-range, when calculated in Hz, but not when defined in semitones. Listeners' evaluation and speakers' self-evaluation of voice-femininity were significantly correlated. However, TVQ^{MtF} scores were significantly correlated only with the speakers' voice-femininity ratings, but not with those of the listeners.

Conclusion: Higher F0 and F1, which are perceived as more feminine, jointly improved speakers' satisfaction with their voice. Speakers' self-evaluation of voice-femininity does not mirror listeners' judgment, as it is affected by additional factors, related to self-satisfaction and personal experience. Combining listeners' and speakers' voice evaluation with acoustic analysis is valuable by providing a more holistic view on how transgender women feel about their voice and how it is perceived by listeners.

KEYWORDS
transgender; women; acoustic analysis; voice therapy; TVQMtF; Hebrew

Introduction

Vocalization is a basic manner by which people communicate and portray their identity and is considered one of the important secondary gender characteristics. Therefore, a gender-congruent voice, in which the individual's voice corresponds with the gender identity, is crucial to one's well-being and quality-of-life (Davies & Goldberg, 2006). Transgender people often show a gender non-conforming voice, as a result of a discrepancy between their gender identity and their perceived gender. Consequently, transgender women who wish for their voice to be perceived in accordance with their feminine identity could be perceived as masculine, and experience gender dysphoria. In a study that surveyed 28 transgender women, over 50% reported viewing the feminization of their communication highly important for their well-being (Neumann et al., 2002). This was further supported by other studies that have demonstrated the importance of feminine voice for transgender women, and how specific voice properties may contribute to the perception of transgender women's voices as feminine (Davies et al., 2015; Stewart, Oates, & O'Halloran, 2020).

In a systematic review, many acoustic measures were examined in association with the perception of speaker's gender (Leung, Oates, & Chan,

CONTACT Ofer Amir oferamir@tauex.tau.ac.il Department of Communication Disorders, Tel-Aviv University, Sheba Medical Center, 52621, Israel.

2018). The two primary acoustic measures consistently shown to differentiate between genders are (a) fundamental frequency (F0), which is perceived as pitch, and (b) formant frequencies, which are a representation of the vocal resonance. In general, mean F0 values range between 100 and 146Hz for cisgender men, and between 188 and 221 Hz for cisgender women (Baken, 1987). Formant frequencies are typically approximately 20% higher for women than for men (Coleman, 1983), with specific normative values reported for English (Peterson & Barney, 1952), for Hebrew (Most et al., 2000), and for many other languages (e.g., Kent & Vorperian, 2018).

Raising F0 typically facilitates the perception of voice as more feminine. Accordingly, transgender women who vocalize with F0 values which are closer to the typical cisgender female range are perceived as more feminine (Owen & Hancock, 2010; McNeill, Wilson, Clark, & Deaking, 2008). Nonetheless, although F0 is the most studied acoustic measure in voice research among transgender women, researchers disagree on the threshold required for differentiating between masculine and feminine voices. On the one hand, a mean F0 value of 155-180Hz is typically considered a virtual "boundary" between the voices of men and women (Davies et al., 2015). But on the other hand, it is clear that higher F0 is not the sole marker that governs the perception of a voice as feminine (Gorham-Rowan & Morris, 2006; Mount & Salmon, 1988). Likewise, the voice of many transgender women who exhibit higher F0, well above the typical masculine range, may still be perceived as non-feminine (Childers & Wu, 1991; Coleman, 1983; Gelfer & Schofield, 2000). This is often attributed to gender differences in formant frequencies; and it was shown that indeed training transgender women to raise formant frequencies facilitates the perception of their voice as more feminine (Carew et al., 2007; Kawitzky & McAllister, 2020).

Intonation is a supra-segmental feature, which was also suggested as a vocal gender marker. Greater variability in F0 during speech and raising intonation contours are considered a feminine trait, whereas a narrower F0 range and "flat" intonation contours are considered more masculine (Graddol & Swann, 1983; Pickering et al., 2012; Stoicheff, 1981). Nonetheless, studies that examined gender differences in vocal-range using acoustic analysis have yielded contradictory results. These differences were attributed to various methodological issues, and mainly to the metric used for quantifying vocal-range. Specifically, vocal-range can be represented in Hertz, which is the physical scale for quantifying the vibratory rate of the vocal mechanism. Yet, vocal-range can also be represented in semitones, which is a perceptual logarithmic scale that better represents the manner in which the human auditory system perceives pitch (Graddol, 1986; Henton, 1989; Nolan, 2003).

Research of vocal-range and intonation contours among transgender women has yielded inconsistent findings. Gelfer and Schofield (2000), for example, reported that transgender women whose voices were judged as feminine had higher values of both lowest and highest F0. The same speakers have also used raising intonation contours more frequently than transgender women whose voices were not judged as feminine. In contrast, Owen and Hancock (2010) reported that transgender women who used a narrower vocal-range were perceived as more feminine. This was later refuted, in a more recent study, that found a wider vocal-range and raising intonation contours among speakers who were judged as more feminine (Hancock et al., 2014). It should be noted that the different studies used different speaking tasks (speech, reading or picture description). This could have contributed to the divergences in the results, as each speech task elicits different intonation patterns and vocal-range. Due to these inconsistencies, and in light of the importance of these features, and the need for studying this topic among transgender women in different cultures (Davies et al., 2015; Hancock et al., 2014); it was deemed necessary to examine 'vocal-range', as one of the acoustic measures in this study, that probed—for the first time—Hebrew speaking transgender women.

The individual's beliefs and perception of their own voice and the impact of that on daily living are an important perspective within the clinical evaluation of voice (Dacakis, Davis, Oates, Douglas, & Johnson, 2013). Transgender women often view their voice as inappropriate for their gender. This could impact their

emotional, social and vocational engagement, and might lead to avoidance behaviors and to a reduced quality-of-life (Davis & Johnson, 2015; Oates & Dacakis, 2015; Stewart, Oates, & O'Halloran, 2020). Considering these aspects is of great interest both socially and professionally, because direct evaluation of the voice alone does not necessarily reflect on how it affects the speakers' life (Geneid et al., 2015; Remacle et al., 2011). Acknowledging the unique experiences and challenges faced by transgender women with regard to their voice has led to the initial development of the TSEQ (Transgender Self-Evaluation Questionnaire) (Davies & Goldberg, 2006); and later to the presentation of the more specific TVQ^{MtF} (Transsexual Voice Questionnaire: Male-to-Female) (Dacakis, Davis, Oates, Douglas, & Johnson, 2013). The TVQ^{MtF} was shown to have strong internal consistency and reliability (Dacakis et al., 2017a, 2017b), and to provide a comprehensive and reliable representation of the way transgender women feel about their voice and its impact on their daily life. Yet, the validity and reliability of the TVQ^{MtF}, and especially the appropriateness of its organization into various number of separate factors is still studied in different languages (Bultynck, Pas, Defreyne, Cosyns, & T'Sjoen, 2020; Salm et al., 2020).

The association between quality-of-life and self-perception of voice-femininity among transgender women has been examined in previous studies. McNeill et al. (2008), for example, reported a strong correlation between self-satisfaction with the voice (on a 'happy-unhappy' scale) and the perception of the voice as feminine. In contrast, in the same study, F0 values were not significantly associated with the speakers' self-satisfaction. It was, therefore, suggested that other vocal properties, apart from F0, influence the speaker's self-satisfaction with her voice. The association between voice-related quality-of-life and the perception of the voice as feminine was later supported by other studies that used the TSEQ (Hancock et al., 2011) and the TVQ^{MtF} (Dacakis et al., 2017a). These studies suggested that quality-of-life among transgender women could be related to their subjective self-perception, but not necessarily or directly to listeners' perception of voice-femininity. Nonetheless, the association between voice-femininity, voice-related quality-of-life and specific voice properties among transgender women is still unclear, and warrants further examination (Dacakis et al., 2017a).

In light of the contradictions between previous reports, the present study examined how specific acoustic voice properties (F0, formant frequencies, vocal-range) are associated with self-perception of voice-femininity among transgender women and with listeners' perception. We also examined how these measures are associated with the speakers' voice-related quality-of-life, as quantified by the TVQ^{MtF}.

Methods

This study was conducted after obtaining the approval of the Tel-Aviv University Ethics Committee, and after all participants had completed and signed an informed consent form.

Speakers

Thirty transgender women, age range 18–42 years ($M = 28.5$, $SD = 6.1$), volunteered to participate in the study. All were native Hebrew speakers, with no reported history of speech, voice or hearing problems. They were recruited through ads on social media and through personal encounters. Four additional potential speakers were excluded due to failing to meet the inclusion criteria. Twenty-four of the speakers reported living in their feminine identity continuously and consistently, four reported living in their feminine identity most of their days, and the remaining two reported that for about 50% of the time. Of the 30 speakers, 21 reported being employed on a permanent position. The reported time elapsed from the event they addressed as the beginning of their gender transition ranged between 1 and 18 years ($M = 5.7$, $SD = 4.2$). Twenty-eight speakers reported receiving hormone therapy for a period ranging between 1 and 18 years ($M = 4.0$, $SD = 4.2$), and 15 used medications for physical and/or psychiatric conditions. Twenty speakers reported practicing their voice with a speech pathologist; ten of them in a group setting, one in an individual setting, one in the form of singing lessons and the remaining eight reported

practicing independently using freely available information on the internet.

Listeners

Ten cisgender men and ten cisgender women (mean age 32.4 years, SD = 4.1) volunteered for the listening task, after responding to ads on local social media forums. All listeners were native Hebrew speakers, with no reported speech or hearing problems. Previous research has suggested that listener's sexual orientation could affect the perception of voice-femininity (Hancock & Pool, 2017). Hence, to reduce the possibility of sexual orientation bias, only listeners who had reported themselves as heterosexual were included. Five of the listeners reported no familiarity with transgender women, nor with transgender voice; five reported infrequent exposures through public media, eight reported distant acquaintances, and two reported familiarity with transgender women. To reduce a possible bias due to revealing the purpose of the study, the anamnesis questionnaire was completed by the listeners only after they had completed the listening task.

Recordings

Each speaker was recorded individually in a quiet room, while performing the following tasks: (a) sustaining the vowel/a/six times for 3–5 seconds, (b) reading two paragraphs from the phonemically balanced reading-passage "The Thousand Islands" (Amir & Levine-Yundof, 2013), and (c) talking on a given topic for five minutes. Digital audio recordings were performed using an Audio-Technica BP892cW-TH MicroSet® headset microphone, located 5 cm from the speaker's mouth, connected to a Centrance pre-amplifier, onto the Audacity® software (ver. 2.1.2), with a sampling rate of 48 kHz (16Bit). Recordings were normalized for amplitude and for DC offset zero crossing. Speakers were instructed to vocalize and speak at ease and no demonstrations were given. The three tasks were recorded in a random order that was modified between speakers. Individual recording sessions lasted approximately 30 minutes.

Acoustic analyses

Acoustic analyses were performed using Praat software (ver. 6.0.10) (Boersma & Weenink, 2019), after manual inspection of the signal and correction of F0 octave-errors. From the recordings obtained from each speaker, six 1-second segments were manually extracted for analysis from the relatively stable mid-section (which may be referred to as the "steady state") of the isolated/a/vowels. All 12 sentences obtained from the reading task were analyzed, as well as 20 phrases from each speaker's free speech. For this purpose, a phrase was defined as a string of words bounded by a single intonation contour, containing a minimum of three words or five syllables (Rochman & Amir, 2013).

From the recordings of the isolated vowels, a mean F0 value was calculated for each speaker, and mean values of the first three formants (F1, F2, F3). Formant frequencies were extracted manually, using the Praat program, after visual inspection of each segment, to avoid measurement errors. A mean F0 value was also calculated for each speaker from the recordings of the reading and speech tasks. Measurements were taken from a combined view of the spectrogram and the F0 tracking display. In addition, each measurement was evaluated perceptually (i.e., auditorily). Special attention was given to instances of vocal fry. Hence, vocal fry frequencies were included in the analyses, after ensuring that the acoustic analysis represents the perceived pitch. F0-range was calculated by subtracting the minimum-F0 value from the maximum-F0 value. This was first calculated for each utterance individually, and then a mean value was derived for each speaker within each task. Vocal-range was calculated and presented in Hertz (Hz.), as well as in semitones (ST).

Listening test of voice-femininity

For the listeners' perceptual task, three sustained productions of the vowel/a/and three sentences from the reading task of each speaker were selected randomly, yielding a total of 180 stimuli (90 vowels + 90 sentences). Samples from the free-speech task were not included in the listening task. This was decided because recordings

contained contextual gender markers that could bias listeners' judgment. Moreover, gender encoding classification is an inherent grammatical property of Hebrew morphology, affecting nouns, verbs and adjectives (e.g., Ravid, 2012).

Recordings were presented individually to the ten listeners, using a SuperLab program, through Sony MDR-CD380 headphones, at a comfortable hearing level set by the listeners independently. To reduce a possible order effect, recordings were presented in a random order that was modified between listeners. Each stimulus was played once, and the listeners were required to rate voice-femininity on a 7-point scale, on which the two extremes were labeled 'masculine' and 'feminine', with no intermediate labels. Ten sustained vowels and ten sentences were randomly selected and presented again to the listeners, for intra-judge reliability evaluation. Consequently, each listener rated a total of 200 recordings (90 vowels + 10 reliability items + 90 sentences + 10 reliability items). Listeners were not given a time limitation for completing the task, but each stimulus was presented once, and ratings could not be modified once made. Mean duration of the individual listening task was approximately 20 minutes. Listeners were allowed to take short recesses during the task, but all chose to perform it continuously.

Intra- and inter-judge reliability

Intra-judge reliability was evaluated using a Pearson correlation test, that confirmed a strong correlation between the repeated judgments ($r = 0.858$, $p < .001$). In addition, strong inter-judge agreement was found, using Cronbach's alpha coefficient ($\alpha = 0.98$). Finally, the possibility of listeners' gender differences in the perception of voice-femininity was tested using two separate Pearson correlation tests, one for vowels and one for sentences. Results confirmed strong and significant correlations between the ratings made by male and female listeners for both vowels ($r = 0.95$, $p < .001$) and sentences ($r = 0.96$, $p < .001$).

Self-perception of voice

Each speaker rated her voice on the same 7-point scale as was used by the listeners, ranging from 'masculine' to 'feminine'. In addition, speakers were asked to rate their satisfaction with their voice, on an un-numbered 7-point scale, ranging from "very satisfied" to "very dissatisfied". Similar rating scales were used previously in studies that employed similar paradigms (Carew et al., 2007; McNeill et al. 2008), hence were regarded appropriate for this task.

Transsexual Voice Questionnaire (TVQ^{MtF}) —Hebrew version

All speakers completed the Hebrew version of the TVQ^{MtF} (Amir & Diamant, 2016), as a measure of voice-related quality-of-life. The total score of the questionnaire, based on all 30 items was used in this study, as a single measure.

Statistical analyses

Statistical analyses were performed using SPSS Ver. 25 (IBM©, SPSS©, 2017). The research variables were described using means and standard deviations. Univariate relations between research variables were expressed as Pearson correlation coefficients, and multivariate relations were estimated within linear regression models. To avoid inflation of Type I error, an FDR correction was used (Benjamini & Hochberg, 1995), with experiment-wise error set at .10.

Results

Descriptive statistics for the acoustic measures obtained from the speakers in the prolonged vowel production, in the reading and speech tasks are presented in Table 1. As shown, similar F0 values were obtained in all three tasks. In addition, the observed vocal-range was slightly larger during the reading task, compared to the speech task.

A summary of the descriptive statistics for the listeners' responses on the voice-femininity scale, as well as the speakers' responses on the voice-femininity scale, the TVQ^{MtF} questionnaire and on the self-satisfaction scale are presented in Table 2.

Association between listeners' perception of voice-femininity and acoustic measures

Table 3 presents a summary of the correlation coefficients obtained between the acoustic

Table 1. Group means, standard deviations and range of values for the acoustic measures obtained in the different speech/voice recording tasks.

Measures	Task	M	SD	Min.	Max.
Mean F0 (Hz.)	Prolonged /a/	155	36	91	248
	Reading	151	33	108	251
	Speech	148	32	103	248
F1 (Hz.)	Prolonged /a/	696	139	356	912
F2 (Hz.)	Prolonged /a/	1387	132	1122	1637
F3 (Hz.)	Prolonged /a/	2619	360	1950	3455
Min. F0 (Hz.)	Reading	106	28	72	205
	Speech	110	27	75	204
Max F0 (Hz.)	Reading	227	53	150	384
	Speech	219	48	157	385
F0-range (Hz.)	Reading	120	36	48	195
	Speech	110	32.	52	181
F0-range (ST)	Reading	13.24	3.23	6.66	22.61
	Speech	11.97	2.94	6.88	19.44

Hz.—Hertz.
ST—Semi tones.

measures of the recordings included in the listening task and the listeners' responses on the voice-femininity scale, as well as the responses of the speakers on the three scales (voice-femininity scale, TVQ^{MtF} and self-satisfaction scale).

The data presented in Table 3 show strong and significant correlations between most acoustic measures and the perception of voice-femininity by listeners. Specifically, listeners' perception of the speakers' voice-femininity was strongly correlated with the speakers' fundamental frequency (F0) on both the vowel and reading tasks. In other words, voices with higher F0 were perceived by listeners as more feminine. Positive correlation coefficients were also found between the values of the first two formants (F1, F2) and the perception of the voice-femininity by listeners. In contrast, values of the third formant (F3) were not significantly correlated with the perception of voice-femininity.

Most acoustic measures related to vocal-range were also significantly correlated with the perception of voice-femininity by listeners. Specifically, speakers who produced voices with higher values of minimum and maximum F0s were perceived as more feminine. Vocal-range, per se (i.e., the difference between maximum and minimum F0 values), was also significantly correlated with listeners' perception of voice-femininity, such that speakers with a wider vocal-range were rated as more feminine. As shown, after converting the F0-range value from the linear Hz scale to the logarithmic Semitone (ST) scale, which is more equivalent to the way the human auditory system perceives pitch, the correlation between the vocal-range measure and the listeners' perception has failed to reach statistical significance.

Table 2. Responses obtained from the listeners on the perceptual voice-femininity scale, and responses obtained from the speakers on the voice-femininity scale, the TVQ^{MtF} questionnaire and the self-satisfaction scale.

Measure	*N*	*M*	SD	Min.	Max.
Listeners—Voice-femininity scale	20	3.28	1.47	1.47	6.53
Speakers—Voice-femininity scale	30	3.87	1.53	2.00	7.00
Speakers—TVQ^{MtF}—total Score	30	63.26	21.35	30.00	96.00
Speakers—Self-satisfaction scale	30	3.83	1.66	1.00	7.00

Notes: Voice-femininity was rated on a 7-points scale (1 = Masculine, 7 = Feminine). Self-satisfaction was rated on 7-points scale (1 = Very unsatisfied, 7 = Very satisfied).

Finally, a multiple regression analysis was performed to predict listeners' perception of voice-femininity, based on the acoustic measures. All acoustic measures that significantly correlated with the listeners' perception (see Table 3) were included simultaneously in the model. Four of these measures contributed significantly to the model, with $R^2 = 0.88$, $F_{(4,25)} = 44.58$, $p < .0001$. These measures were F0, F1 and F2 of the vowel/a/and F0 extracted from the reading task. Their standardized parameters (betas) ranged between .22 and .33, with all p's $< .03$.

Association between speakers' scales and acoustic measures

Table 3 presents a summary of the correlation coefficients obtained between all acoustic measures and the subjective scales rated by the speakers. As noted above, the speakers completed three scales, from which three measures were extracted. These included the voice-femininity scale, the TVQ^{MtF} total score and the self-satisfaction scale. A summary of the results follows, for the association between each of the three measures and the acoustic measures examined.

Self-perception of voice-femininity and acoustic measures

Inspection of the correlation coefficients obtained for the *speakers'* voice-femininity scale (Table 3) presents similar results to those obtained from the equivalent *listeners'* scale, with only minor differences. In general, most examined acoustic measures were significantly correlated with the self-perception of voice-femininity. Specifically,

Table 3. Summary of Pearson correlation coefficients obtained between the listeners' and speakers' scales and the acoustic measures extracted from the recordings.

Measures	Listeners	Speakers		
	Voice-femininity scale	Voice-femininity scale	TVQ^{MtF}	Self-satisfaction
Mean F0—/a/	0.84**	0.64**	−0.39*	0.41*
F0—Reading	0.84**	0.61**	−0.39*	0.37*
F1—/a/	0.64**	0.49**	−0.13	0.26
F2—/a/	0.71*	0.25	0.01	−0.02
F3—/a/	0.23	0.31	−0.05	0.20
Min. F0 – Reading	0.73**	0.54**	−0.36*	0.31
Max. F0 – Reading	0.77**	0.63**	−0.44*	0.41*
Range – Reading (Hz)	0.56**	0.50**	−0.36	0.36*
Range—Reading (ST)	0.01	0.12	−0.01	0.08

*$p < 0.05$; **$p < 0.01$.

Notes: Voice-femininity was rated on a 7-points scale (1 = Masculine, 7 = Feminine). Self-satisfaction was rated on 7-points scale (1 = Very unsatisfied, 7 = Very satisfied).

Table 4. Correlation coefficients between listeners' and speakers' responses on the subjective voice-femininity scales, TVQ^{MtF} and self-satisfaction scale.

Group	Scales	Speakers		
		Voice-femininity scale	TVQ^{MtF}	Self-satisfaction
Listeners	Voice-femininity scale	0.59**	−0.25	0.29
Speakers	Voice-femininity scale	–	−0.74**	0.78**
	TVQ^{MtF}	–	–	0.80**

**$p < 0.01$.

self-perception of voice-femininity was moderately correlated with the two F0 measures. A positive correlation coefficient was also found between F1 values and the self-perception of voice-femininity. However, both F2 and F3 have yielded weak and insignificant correlation coefficients with the self-perception of voice-femininity.

Similarly, most vocal-range measures were significantly correlated with the speakers' self-perception of voice-femininity. Namely, speakers with higher values of minimum and maximum F0s have rated their voice as more feminine. Vocal-range on the reading task was also significantly correlated with self-perception of voice-femininity. In addition, similar to the results obtained from the listeners' scale, converting the vocal-range scale from Hz to Semitones has eradicated the significant correlations.

A multiple regression analysis was performed to predict self-perception of voice-femininity based on the acoustic measures, as performed above. Results demonstrated that F0 of the vowel/a/was the only predictor for self-perception of voice-femininity, with $R^2 = 0.41$, $F_{(1,26)} = 18.38$, $p < .0001$.

TVQ^{MtF} and acoustic measures

Data in Table 3 show negative correlation coefficients between the F0 measures and the TVQ^{MtF} scores. In other words, speakers with a higher F0 reported more positively on their voice and on their voice-related experiences. Similarly, minimum and maximum F0 values during the reading task were also negatively correlated with the TVQ^{MtF} scores. All other acoustic measures did not significantly correlate with the TVQ^{MtF} scores. A regression analysis performed on this set of data revealed that maximum F0 in the reading task was the only measure that predicted TVQ^{MtF} scores [$R^2 = 0.19$, $F_{(1,28)} = 33.23$, $p < .05$].

Voice-femininity scales and TVQ^{MtF}

Table 4 shows strong and significant correlation between speakers' self-evaluation of voice-femininity and the TVQ^{MtF} scores.

In other words, speakers who perceived their voice as more feminine have exhibited lower scores on the TVQ^{MtF}. Similarly, they have also reported being more satisfied with their voice. On the other hand, no significant correlation was found between the *listeners'* evaluation of voice-femininity and the *speakers'* responses on the TVQ^{MtF} nor on the self-satisfaction scale.

Finally, a moderate correlation was found between the listeners' and speakers' voice-femininity scales. That is, speakers who perceived their

voice as more feminine were perceived similarly by the listeners.

Discussion

Voice characteristics and the perception of voice-femininity

Fundamental frequency (F0) and formant frequencies

Our findings indicate strong and significant correlations between all measures of fundamental frequency (F0) and the perception of voice-femininity by the listeners and by the speakers. This is in agreement with the general view of F0 as a key feature in the perception of voice-gender (e.g., McNeill et al., 2008; Owen & Hancock, 2010). Interestingly, the correlation between F0 and *listeners'* judgment was strong (r=.84), whereas the correlation between F0 and *speakers'* self-judgment of their voice was moderate (.54≤r≤.64). In other words, both listeners and speakers have associated higher F0 values with feminine voice. Nonetheless, while the transgender women in our study have, indeed, associated their pitch with their perception of voice-femininity, they appear to have associated the two factors less strongly than did the listeners. This is interpreted as supporting previous research that suggested that *self*-perception of voice femininity *does not* mirror listeners' perception, as it relies on additional factors (i.e., Owen & Hancock, 2010).

The clinical implication of this finding should be considered with caution, though, because raising pitch per se might not be sufficient as an independent and ultimate goal for voice therapy. Mount and Salmon (1988), for example, have demonstrated that a transgender woman who successfully raised her F0 to 210 Hz (i.e., well within the typical feminine range), was still not perceived as portraying a feminine voice. Only after focusing speech-therapy on altering tongue positioning and modifying vocal resonance, did her voice quality shift and she was judged as feminine in face-to-face interactions as well as over the phone. Similar findings were reported by others (Coleman, 1983; Gelfer & Schofield, 2000; Kawitzky & McAllister, 2020), suggesting that raising F0 alone in transgender women indeed facilitates the perception of voice-femininity, but requires additional voice modifications.

When examining formant frequencies independently of F0, values of the first two formants (F1 and F2) were significantly correlated with the perception of voice-femininity by listeners. Speakers' perception, however, was correlated with F1, but not with F2 or F3. For a more holistic examination of this question, and to better simulate the way listeners perceive speakers' voice, a multiple regression analysis was performed, combining all significant acoustic measures concurrently. This analysis has shown that speakers' self-perception of voice-femininity was predicted by F0 alone. *Listeners'* perception, on the other hand, could be predicted by the combined effect of F0 and the first two formants (F1 and F2). This is reminiscent of previous reports that have highlighted the combined effect of F0 and formant frequencies on the perception of voice-gender (Gelfer & Mikos, 2005; Pasricha et al., 2008). It should be noted that the combined effect of F0 and formant frequencies on the perception of voice-gender was not only shown in transgender speakers, but also in cisgender women and men, using computerized (i.e., synthesized) modifications. Hillenbrand and Clark (2009), for example, have shown that modifying either F0 or formant frequencies artificially, to fit the typical ranges of the opposite gender, was mostly ineffective; and did not lead to a perceptual change in voice-gender.

When considering the importance of F0 and formant frequencies for the perception of voice-gender, a highly relevant topic is the perceptual *threshold* that should be reached for the voice of transgender women to be perceived is feminine. Admittedly, this discussion is beyond the scope of our study, as the current methodology was not designed to address this question. Nevertheless, a qualitative inspection of our raw data demonstrates that all speakers whose voices were rated as feminine (i.e., higher than 4.5 on the voice femininity 7-point scale) had F0 values of 168 Hz or higher. In contrast, all speakers whose voices were rated as masculine (i.e., lower than 3.5 on the voice femininity scale) had lower F0 values than 161 Hz. This observation supports previous

reports that suggested a minimum F0 value of 155-180Hz as a prerequisite for a transgender voice to be perceived as feminine (Davies et al., 2015). Interestingly, while our data portray a rather dichotomous distribution of F0 values, it does not show a similar pattern for formant frequencies. In other words, formant frequencies alone did not separate speakers' voices on the perceptual masculine-feminine scale.

Vocal-range: F0 and semitones

In light of the theoretical debate on the appropriate metric for quantifying vocal-range (e.g., Graddol, 1986; Henton, 1989), both the linear (Hz.) scale and the perceptual semitone (ST) scale were used here. Our results demonstrated that, using the Hz scale, speakers' vocal-range was correlated with both listeners' and speakers' perception of voice-femininity. However, transformation to the ST scale has eliminated these findings, yielding no significant correlations. That is, speakers who used a wider range of F0 values were perceived as having a more feminine voice. But when data were examined in semitones, no association was found between vocal-range and voice-gender. This seeming paradox can be resolved in light of the significant correlations between the nominal values of minimum- and maximum-F0 and the voice-femininity judgments. As shown, speakers with higher values of both minimum- and maximum-F0 were perceived as having a more feminine voice. Therefore, due to the logarithmic nature of the ST scale, a speaker with elevated F0 is more likely to exhibit a wider vocal-range, when calculated in Hz.

A similar conclusion was reached by Henton (1989), who reviewed a body of studies that reported data using the Hz scale, and converted their reports to the ST scale. She concluded that the conception of a wider vocal-range in women may not be, in fact, accurate. It should be noted that this discussion is not yet settled, due to methodological differences between studies that examined this issue among cis- and trans-gender speakers in various speaking/reading conditions (Gelfer & Schofield, 2000; Owen & Hancock, 2010). Therefore, it is suggested that when a transgender women aim for a more feminine voice, she would benefit from raising her minimum- and maximum-pitch, which will increase the eventual F0-range. This is expected to affect listeners' perception, even if the perceptual vocal-range (i.e., using ST scale) is not significantly widened. The current data is not sufficient to determine whether the upper or the lower limit of the vocal-range is more influential in the perception of voice-femininity. Yet, our clinical experience suggests that transgender women whose minimum-F0 is markedly low are typically perceived as portraying a more masculine voice altogether. At the same time, those with higher maximum-F0 are not necessarily perceived as more feminine. Hence, we suggest that, when targeting pitch in voice-therapy, elevating minimum-pitch could be more effective clinically than elevating maximum-pitch. Nonetheless, this should be examined more directly in future research. In addition, one must not overlook the fact that our data are mostly based on a reading task, because it was imperative to ensure that no linguistic gender markers are evident in the recording. Therefore, it is conceivable that different results could be found, if data are obtained from conversational speech.

Our findings highlight the importance of quantifying vocal-range in ST, and not relying on the Hz scale alone. Yet, we demonstrate that combining both scales for examining the same set of data reveals valuable information and provides additional insight that facilitates deeper understanding on how listeners perceive voice-femininity. We suggest that this combination of metrics should be used in future studies, as it could provide additional insight, as well as improve efficacy and focus of voice-therapy provided to transgender women.

TVQ^{MtF} scores and acoustic measures

The present results demonstrate that speakers who scored lower on the TVQ^{MtF} have also exhibited higher values of all F0 measures. However, the regression analysis performed on this set of data has shown that maximum-F0 was the single acoustic measure that significantly predicted the speakers' TVQ^{MtF} scores. This result is of interest, as it demonstrates that the upper boundary of the vocal-range could predict, limitedly though, how transgender women feel about

their voice and about their daily experience with it. While this finding should be replicated before a solid conclusion may be drawn from it, this may be interpreted as demonstrating that voice-femininity is not equivalent to the subjective experience of transgender women regarding their voice. That is, higher ratings of voice-femininity do not necessarily guaranty high self-satisfaction with one's voice; and vice-versa, lower ratings of voice-femininity does not necessarily lead to low self-satisfaction with one's voice.

Evaluating this result in light of previous publications is challenging, as only a few studies have entertained the possibility of an association between specific acoustic measures and the subjective evaluation of quality-of-life or self-satisfaction. Dacakis et al. (2017a), for example, explored the association between TVQ^{MtF} and F0 in 34 transgender women, and reported no significant correlations. They, too, acknowledged the scarce data on this topic, and suggested that vocal pitch and voice-related quality-of-life may not be directly related. A similar conclusion was also drawn by McNeill et al. (2008), who reported no significant correlation between speakers' F0 and their "satisfaction with self-perceived femininity of the voice" in a group of 12 transgender women.

Nonetheless, our data show significant correlations between the TVQ^{MtF} scores and all F0 measures. These findings are supported by the significant associations between the acoustic measures and the speakers' responses on the voice-femininity scale, as well as on the self-satisfaction scale. Hence, it is suggested that this association should not be dismissed, and should be further examined in future studies.

TVQ^{MtF} scores and the perception of voice-femininity

The present study combined speakers' and listeners' evaluations, thus it provides the opportunity of combining and comparing both views. Data show a strong and significant correlation between speakers' self-perception of their voice-femininity and their scores on the TVQ^{MtF} (see Table 4). However, no such significant correlations were found with *listeners'* perception of the speakers' voice-femininity. This finding is of special interest in light of the significant correlation between the listeners' and speakers' responses on the voice-femininity score. In other words, although the ratings of voice-femininity made by both speakers and listeners were in general agreement, only the speakers' ratings correlated with the TVQ^{MtF} scores. This supports the efficacy of the TVQ^{MtF} in portraying the unique internal perspective of the transgender women about their voice, independently of the listener's perspective.

This result also supports previous reports on a strong association between the TVQ^{MtF} scores and self-perception of voice-femininity (Dacakis et al., 2017a), but on much weaker associations between the TVQ^{MtF} scores and listeners' perception of speakers' voice-femininity (Owen & Hancock, 2010). This is an empirical representation of the complexity of gender perception.

Study limitations

Two methodological caveats of the current study should be noted. First, our sample was based on volunteers, and therefore it might not provide a balanced representation of the transgender women population in Israel. It may be argued that those who volunteered for participation could have experienced more difficulties with their voice, hence seek professional assistance more than those who are more satisfied with their voice. Second, this study evaluated voice-femininity using a 7-point rating scale, rather than a binary (male:female) scale. This taps into the important discussion of whether gender is generally judged as a binary state or on a continuum. Therefore, it may be argued that different results could have been obtained if listeners and speakers in our study were asked to evaluate voice-femininity differently.

Conclusion

Findings demonstrated that measures of F0 as well as formant frequencies were associated with the perception of voice-femininity. It also revealed a significant association between F0, and especially maximum-F0, with voice-related quality-of life. In addition, data demonstrated that listeners and speakers evaluated voice-femininity

differently, such that speakers' judgment of their voice-femininity was less strictly associated with their pitch, in comparison with the listeners.

Acknowledgments

The authors wish to thank the participants for their openness and contribution to this study. Dr. Ilan Roziner is greatly acknowledged for his assistance and guidance in the performance of the data analyses. The Màavarim ("Passageways") Association is acknowledged for the assistance in approaching and recruiting potential participants. Dr. Noam Amir is acknowledged for his help in computerizing the listening task.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

ORCID

Ofer Amir http://orcid.org/0000-0002-2873-954X

References

Most, T., Amir, O., & Tobin, Y. (2000). The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech*, *43*(Pt 3), 295–308. https://doi.org/10.1177/00238309000430030401

Amir, O., & Diamant, N. (2016). TVQ^{MtF} Authorized Hebrew version. Shelagh Davies website. http://www.shelaghdavies.com/questionnaire/hebrew_authorised_translation.pdf.

Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of Voice*, *27*(4), 524.e1–524.e10. https://doi.org/10.1016/j.jvoice.2013.01.015

Baken, R. J. (1987). *Clinical measurement of speech and voice*. Brown and Company, Inc.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological))*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Ver. 6.1.05 http://www.praat.org.

Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., & T'Sjoen, G. (2020). Organizing the voice questionnaire for transgender persons. *International Journal of Transgender Health*, *21*(1), 89–97. https://doi.org/10.1080/15532739.2019.1605555

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

Childers, D., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *The Journal of the Acoustical Society of America*, *90*(4 Pt 1), 1841–1856. https://doi.org/10.1121/1.401664

Coleman, R. (1983). Acoustic correlates of speaker sex identification: Implications for the transsexual voice. *The Journal of Sex Research*, *19*(3), 293–295. https://doi.org/10.1080/00224498309551189

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

Dacakis, G., Oates, J., & Douglas, J. (2017a). Associations between the Transsexual Voice Questionnaire (TVQMtF) and self-report of voice femininity and acoustic voice measures . *International Journal of Language & Communication Disorders*, *52*(6), 831–838. https://doi.org/10.1111/1460-6984.12319

Dacakis, G., Oates, J., & Douglas, J. (2017b). Further evidence of the construct validity of the Transsexual Voice Questionnaire (TVQMtF) using principal components analysis. *Journal of Voice : official Journal of the Voice Foundation*, *31*(2), 142–148. https://doi.org/10.1016/j.jvoice.2016.07.001

Davies, S., & Goldberg, J. (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., & 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

Davis, S., & Johnson, J. (2015). Exploring the validity of the Transsexual Voice Questionnaire for male-to-female transsexuals. *Revue Canadienne d'Orthophnie et d'Audiologie*, *39*(1), 40–51.

Gelfer, M., & Mikos, V. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels.

Journal of Voice, *19*(4), 544–554. https://doi.org/10.1016/j.jvoice.2004.10.006

Gelfer, M., & Schofield, K. (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

Geneid, A., Rihkanen, H., & Kinnari, T. (2015). Long-term outcome of endoscopic shortening and stiffening of the vocal folds to raise the pitch. *European Archives of Oto-Rhino-Laryngology : official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology—Head and Neck Surgery*, *272*(12), 3751–3756. https://doi.org/10.1007/s00405-015-3721-7

Gorham-Rowan, M., & Morris, R. (2006). Aerodynamic analysis of male-to-female transgender voice. *Journal of Voice : official Journal of the Voice Foundation*, *20*(2), 251–262. https://doi.org/10.1016/j.jvoice.2005.03.004

Graddol, D. (1986). Discourse specific pitch behaviour. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 221–237). Croom & Helm.

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech*, *26*(4), 351–366. https://doi.org/10.1177/002383098302600403

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. https://doi.org/10.1177/0261927X17704460

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

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

Henton, C. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication*, *9*(4), 299–311. https://doi.org/10.1016/0271-5309(89)90026-8

Hillenbrand, J., & Clark, M. (2009). The role of F0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception & Psychophysics*, *71*(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Kawitzky, D., & McAllister, T. (2020). The effect of formant biofeedback on the feminization of voice in transgender women. *Journal of Voice : Official Journal of the Voice Foundation*, *34*(1), 53–67. https://doi.org/10.1016/j.jvoice.2018.07.017

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, *74*, 74–97. https://doi.org/10.1016/j.jcomdis.2018.05.004

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*, *61*(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067

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

Mount, K., & Salmon, S. (1988). Changing the vocal characteristics of a postoperative transsexual patent: A longitudinal study. *Journal of Communication Disorders*, *21*(3), 229–238. https://doi.org/10.1016/0021-9924(88)90031-7

Neumann, K., Welzel, C., Gonnermann, U., & Wolfradt, U. (2002). Satisfaction of MtF transsexuals with operative voice therapy—a questionnaire-based preliminary study. *International Journal of Transgenderism*, *6*(4), 1–25.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences*. (p. 774). International Phonetic Association.

Oates, J., & Dacakis, G. (2015). Transgender voice and communication: Research evidence underpinning voice intervention for male-to-female transsexual women. *Perspectives on Voice and Voice Disorders*, *25*(2), 48–58. https://doi.org/10.1044/vvd25.2.48

Owen, K., & Hancock, A. (2010). The role of self- and listener-perception of femininity in voice therapy. *International Journal of Transgenderism*, *12*(4), 272–284. https://doi.org/10.1080/15532739.2010.550767

Pasricha, N., Dacakis, G., & Oates, J. (2008). Communicative satisfaction of male-to-female transsexuals. *Logopedics, Phoniatrics, Vocology*, *33*(1), 25–34. https://doi.org/10.1080/14015430701514500

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Pickering, L., Hu, G., & Baker, A. (2012). The pragmatic function of intonation: Cueing agreement and disagreement in spoken English discourse and implications for ELT. In J. Romero-Trilli *Pragmatics and prosody in English language teaching* (pp. 199–218). New York: Springer-Verlag.

Ravid, D. (2012). *Spelling morphology: The psycholinguistics of Hebrew spelling*. Springer.

Remacle, M., Matar, N., Morsomme, D., Veduyckt, I., & Lawson, G. (2011). Glottoplasty for male-to-female transsexualism: Voice results. *Journal of Voice : official Journal of the Voice Foundation*, *25*(1), 120–123. https://doi.org/10.1016/j.jvoice.2009.07.004

Rochman, D., & Amir, O. (2013). Examining in-session expressions of emotions with speech/vocal acoustic measures: An introductory guide. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *23*(4), 381–393. https://doi.org/10.1080/10503307.2013.784421

Salm, S., Hower, K., Neumann, S., & Ansmann, l. (2020). Validation of the German version of the Transsexual Voice Questionnaire for male-to-female transsexual. *Journal of Voice*, *34*(1), 68–77. https://doi.org/10.1016/j.jvoice.2018.06.010

Stewart, L., Oates, J., & O'Halloran, P. (2020). My voice is my identity: The role of voice for trans women's participation in sport. *Journal of Voice*, *34*(1), 78–87. https://doi.org/10.1016/j.jvoice.2018.05.015

Stoicheff, M. (1981). Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech, Language, and Hearing Research*, *24*(3), 437–441. https://doi.org/10.1044/jshr.2403.437