Mindfulness (2023) 14:3069–3085
https://doi.org/10.1007/s12671-023-02264-8

ORIGINAL PAPER

Hindering or Helping? User Preferences for Features of Recorded Mindfulness Training

Maja Stanko-Kaczmarek[1] · Tomasz Jankowski[2] · Anna Koczyk[1] · Wadim Krzyzaniak[1] · Kalina Meder[1] · Kajetan Dulian[1]

Accepted: 1 November 2023 / Published online: 1 December 2023

Abstract

Objectives Despite its numerous benefits, practicing mindfulness involves certain challenges. This study investigates the impact of various characteristics of recorded mindfulness practice instructions on recipients' evaluations, focusing on elements that may evoke negative reactions and hinder mindfulness practice.

Method A total of 138 participants evaluated 24 recordings of mindfulness practice excerpts. The recordings differed in terms of the speaker's gender, the form and pace of the message, the presence of whispering, and the number of artifacts. Respondents rated each recording using a scale and provided open-ended responses about their subjective perception of the recordings. The study employed hierarchical multilevel modeling to analyze the collected data.

Results The results indicated that recipient gender did not influence preferences for male or female-voiced recordings. However, properties such as form, pace, whispering, and artifacts did affect evaluations. Least favored instruction elements included plural pronouns, slow pacing, primarily whispered speech, and a high number of artifacts. Some differences in evaluations were observed between female and male recording conditions. Additionally, qualitative data revealed participants' subjective reactions to recordings with varying characteristics, and the study identified the most favorable characteristics of the recordings.

Conclusions The study results identified which qualities of recorded mindfulness instructions are least preferred and may present obstacles to initiating or continuing the practice. Therefore, this study may help create more optimal instructions and improve the design of apps and platforms offering mindfulness practice recordings, enhancing the quality and accessibility of practice for a broader audience.

Preregistration This study is not preregistered.

Keywords Mindfulness · Obstacles · Instruction · Preferences · Recordings · Voice

Mindfulness has become a flourishing area of research and practice (Baminiwatta & Solangaarachchi, 2021; Goldberg et al., 2017). It entails purposefully focusing one's attention on the present moment while maintaining a non-judgmental and accepting mindset (Brown & Ryan, 2003; Kabat-Zinn, 1994, 2003). Studies have consistently shown that mindfulness is associated with positive outcomes, such as increased levels of life satisfaction, self-esteem, vitality, sense of competence, optimism, and positive emotions (Keng et al., 2011). Conversely, negative correlations have been observed between mindfulness and symptoms of depression, rumination, dissociation, distraction, social anxiety, emotional dysregulation, stress, and experiential avoidance (Tomlinson et al., 2018).

Mindfulness can be developed and strengthened through various forms of independent practice, organized training, and interventions (Germer et al., 2005). In mindfulness practice, instructors typically guide individuals or groups through techniques and exercises that aim to cultivate awareness, attention, and acceptance of present-moment experiences. These exercises may involve assuming a comfortable posture and focusing on the breath or other sensations in the

✉ Maja Stanko-Kaczmarek
m.stanko@amu.edu.pl

[1] The Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, Szamarzewskiego Street 89/AB, 60 – 568 Poznan, Poland

[2] The Institute of Psychology, John Paul II Catholic University of Lublin, Lublin, Poland

body, as well as guided meditations or visualization exercises (Kabat-Zinn, 2013).

Despite the well-documented benefits of mindfulness practice, some individuals may encounter difficulties when trying to practice mindfulness regularly (Farias et al., 2020). Research suggests that individuals may not disclose negative emotions or challenges in practice due to feelings of shame or a desire to please the instructor (Britton et al., 2021). Criticisms or suggestions for changes that could facilitate practice may also be received negatively and ignored (Van Dam et al., 2018). Therefore, it is crucial to examine both the positive effects and challenges associated with mindfulness training (Farias et al., 2020). However, relatively few researchers have focused on exploring the difficulties and challenges associated with mindfulness practice (Germer et al., 2005; Kabat-Zinn, 2017, 2018).

The existing literature indicated obstacles accompanying mindfulness training have addressed practitioner issues, such as limitations in patience, motivation, conscientiousness, and self-discipline (Kabat-Zinn, 2013, 2017). When such difficulties arise, the practitioner is encouraged to solve the problems on their own. For example, one may try changing their mindset, taking more intentional action, or waiting out the moments when there is an urge to quit (Kabat-Zinn, 2013, 2017). Practitioners were also advised that after prolonged practice, the reluctance will spontaneously diminish.

However, in addition to internal barriers, challenges in practicing mindfulness also include issues related to external obstacles. By "external obstacles," we refer to external factors or conditions that may hinder the effectiveness or reception of mindfulness training, such as the instructional feature, procedure, and specific practice forms (Britton et al., 2021). Unfortunately, the instructions' problematic qualities rarely were considered in research (Rothschild, 2017). To date, the literature has only examined these practice elements when practitioners report having previously experienced severe stress and traumatic experiences (Rothschild, 2017; Treleaven, 2018). Attempts to identify difficulties and modify or revise practice instructions or program structure to meet the needs of these practitioners have been made. Nevertheless, analyses concerning the difficulties associated with the formal aspects of mindfulness practice in the general population are still lacking.

According to our observations, many who have participated in structured mindfulness-based training or used ready-made recordings available on the Internet or applications (*apps*) complain about the formal qualities of the instructions. The characteristics of the instructions heard may trigger adverse reactions. Such reactions may be intense enough to make it very difficult or even impossible for the recipient to start, continue, or fully engage and participate in the practice. For example, contact with unpleasant or irritating stimuli in the form of externally administered instructions can cause distress and unpleasant emotions of varying intensity—from anger or annoyance through disgust to feelings of anxiety (Farias et al., 2020; Zhu et al., 2019). Moreover, the elements of instruction evoking such reactions also lead to cognitive difficulties, for example, focusing attention on the instruction content. Thus, practice is significantly impeded, persistence is discouraged, and would-be practitioners abandon the practice. The possibility of reaping benefits from mindfulness is reduced.

In conclusion, despite increasing research on the positive impact and effectiveness of practicing mindfulness, research on the difficulties of its practice remains scarce. Furthermore, data on external factors' impact on hindering mindfulness practice, such as elements in the instruction in mindfulness practice, are insufficient. Therefore, we decided to extend research in this field. Our study aimed to expand and add to our knowledge of difficulties encountered during mindfulness practice. In addition, we test the influence of selected external factors on audience evaluation. More specifically, we wanted to determine which instructional characteristics are perceived as the least desirable and the most poorly evaluated, and may hinder practice in the general population.

Formal properties of heard messages are essential in the response that occurs in humans (McAleer et al., 2014). Even small extratextual nuances affect the listener's emotional response and degree of arousal, as well as their attitude toward the voice, the message content, and its sender (Bhatia et al., 2005; Loui et al., 2013). Recipients have specific preferences concerning the characteristics of voice messages (Ding et al., 2017; Hain et al., 2009). As studies showed, after the first few seconds, the voice is categorized, and certain character traits attributed to its owner. A specific (positive or negative) attitude toward the person emerges, which is dependent on listeners' individual preferences (McAleer et al., 2014). To begin with, the elaborate psychobiological systems that have evolved to recognize and process voices confirm that they significantly impact the recipient (Pisanski & Bryant, 2019).

This remarkable relevance of the voice matters, among other things, in psychotherapy. The appropriate voice serves multiple functions in confidence building, reassuring, or assisting in emotional expression and regulation (Bady, 1985; Blanck et al., 1986). The nature of the listening voice impacts the sense of being understood and noticed (Westland, 2015). It fosters a safe space where the client efficiently opens to therapeutic help (Loewy, 2004).

On the other hand, some voice parameters may be judged undesirable and annoy the recipient. Unfavorably evaluated properties may be, for example, the gender of the voice (Parson et al., 2013). In addition, undesirable reactions in some recipients may also include (a) the grammatical form of messages (including plurals in utterances) and the use of the

imperative mood (Wicklund, 1974), (b) an inappropriate (too fast/slow) speaking rate (Boone et al., 2019), (c) low volume of the spoken message, especially whispering (Murry, 1988), and (d) numerous and audible artifacts (e.g., saliva swallowing, gurgling, breathing) (Cavanna & Seri, 2015).

We have observed people using external instructions to practice mindfulness and exhibiting individual preferences for gendered voices. However, research on gender preference in voice messages is inconclusive. Some study results indicated a general male preference for feminine voices (high-pitched) and a female predilection for masculine voices (lower-pitched). These preferences may have an evolutionary basis (Xu et al., 2013). In contrast, other research reported that in giving instructions, for example, most people prefer a female voice regardless of their gender (Parson et al., 2013). As mindfulness practice instruction takes the form of "directing" the listeners and their attention, we expect that male-voiced messages would not be as well received.

Aspects of the message, such as the grammatical structure and mood of the verb, significantly affect listeners' perception. For instance, the first-person plural form of the message sometimes used in instructions ("we") compared to the second-person singular form ("you") can arouse cognitive dissonance and negative emotions. This effect may be caused by the introduction of a contextually vague sense of connection or intense presence of the trainer in a person's intimate experience, associated with focusing on one's bodily and mental experiences (Rothschild, 2017; Slatcher et al., 2008). Therefore, we expect that a stranger's presence in an individual's intimate space on linguistic grounds may be evaluated unfavorably.

Another factor differentiating recipients' reactions may be the imperative mood in which the instructions are frequently given. This mode may cause greater aversion, compared to the propositional form, due to *reactance*—a distinct sense of discomfort caused by an attempted restriction of individual freedom (Wicklund, 1974). Thus, instructions in mindfulness training given in the imperative mood may be evaluated less favorably compared to those presented in a propositional style. The term "propositional form" refers to a style of instruction that offers suggestions or propositions rather than giving direct commands. For instance, instead of saying "Focus on your breath," a propositional style might say "You could turn your attention to your breathing." On the other hand, mindfulness instructions are typically created in fully or partially imperative mood. Some trainers may already be accustomed to this mode (Rothschild, 2017). Therefore, grammatical form's impact is somewhat ambiguous. However, considering the population, the effect of reactance and aversiveness may be more substantial.

Research showed that a slower instruction pace is associated with monotony (Boone et al., 2019). Thus, slow speech makes it more challenging to focus on the content, including during mindfulness training. In addition, practitioners complain about excessive pauses between sentences co-occurring with a slower speaking pace. These features are rated as distracting and disruptive to the practice rhythm. When the instructor speaks slowly, the listener may pay more attention to the instruction form than its content, which contributes to distraction and practice difficulty (Boone et al., 2019).

The standard procedure when instructing in mindfulness practice is to lower the volume of the voice and deliver all or part of the instruction in a whisper. On the one hand, whispering can engender greater feelings of trust, closeness, and intimacy between the listener and person speaking (Andersen, 2014). On the other hand, some people may judge whispered instructions as crossing the boundaries of formality. Whispering can also evoke unpleasant somatic sensations, such as tingling, which may be found unpleasant, or which may evoke erotic associations (Waldron, 2017). Sexualizing a message due to whispering (even inadvertently) seems to be powerful in messages delivered by women (Hughes et al., 2014). This treatment may repel practitioners due to a lack of contextual appropriateness. Many audiences also report that certain sounds perceived while listening to whispering cause anxiety (Janik McErlean & Banissy, 2018).

A preliminary analysis of mindfulness instructional videos available on free platforms (e.g., YouTube) revealed many auditory artifacts. Auditory artifacts, in this case, are additional sounds accompanying speech, such as audible breathing, mumbling (including gurgling), swallowing saliva, grunts, or other sounds associated with speech production. In recorded instructions, artifacts can be accentuated, becoming more audible and distinct than in a face-to-face situation.

Depending on individual differences, listeners vary in sensitivity to voice-related factors. For example, the artifacts may be related to the growing number of diagnoses of *misophonia* (Janik McErlean & Banissy, 2018). In the condition known as misophonia, small noises, such as mumbling, grunting, coughing, or breathing cause stress, anger, anxiety, and severe agitation (Kumar et al., 2017). Therefore, analysis of general reactions to such artifacts is essential. Determining the reception of messages containing these sounds is crucial from the perspective of design, implementation, and adaptation to listeners' reactions. Making such adjustments will facilitate the reception of instructions for susceptible individuals.

Based on the research results on various properties of voice messages and our observations of the individual reactions using external instructions during mindfulness practice, formal instructional elements are critical for practicing person's reception. They differentiate emerging reactions, attitudes, and the course of practice. This study aimed to determine recipients' preferences for the aforementioned instructional elements. We

addressed the information gap on the reception of mindfulness practice instructions and the inconclusive results from previous studies conducted in other contexts.

The study analyzed the formal features of recorded voice instructions compared to those provided during face-to-face or video practice. It was decided to use audio-only instruction so that subjects' evaluation of message features did not depend on video (e.g., the instructor's appearance). Another rationale for analyzing evaluations of recordings is the current dearth of such research, despite the successful development of technological solutions that enable mindfulness practice, such as playing recorded instructions on mobile devices associated with courses, online platforms, and mobile apps (Chittaro et al., 2016; Economides et al., 2018). We hypothesized that obtaining data on recording properties contributes to the existing literature and supports the preparation of instructions to facilitate engagement in practice for those who use recordings during independent practice, structured training, or research participation.

Our study explored whether different message properties can affect positive or negative evaluation by recipients. Specifically, we sought to determine which recording elements the message listeners least preferred and which voice properties were most preferred. We expected that gender, grammatical form, pace, whispering, and artifacts would influence recording evaluations based on the results of previous studies on other types of messages. More specifically, a difference was anticipated between the ratings based on gender. Evolutionary explanations propose that men prefer messages conveyed by women, and women prefer messages delivered by men. Thus, messages delivered in voice of the listener's gender would receive lower ratings. For the group, we expected that recordings would receive low ratings if they are in the first-person plural and imperative mood, they are delivered at a slow pace, they are primarily whispered, or they feature many artifacts.

Detailed information on the exact characteristics of these types of recordings can be found in the section dedicated to stimulus construction. Another research goal was to explore which recording characteristics are most preferred. The study controlled for the mindfulness practice experience factor, which may have influenced ratings, for example, non-practitioners may judge recordings more harshly than experienced practitioners who are accustomed to some aspects of instruction and would approach them with greater acceptance.

Method

Participants

The study included 138 participants. The majority (51.7%) were aged 18 to 26 years. Table 1 shows the descriptive statistics (the number and percentage distribution of participants by age, gender, and whether they practice mindfulness).

Procedure

The survey was administered using Google Forms by sharing a link with survey participants on social media. At the beginning of the survey, all persons gave their informed consent and provided basic metrics. Next, participants underwent a four-segment procedure involving listening to more recordings, making evaluations, and answering open-ended questions.

Finally, they indicated whether they had performed the survey with headphones, which could influence the differences in the ratings—34.1% of the participants declared that they performed the survey while wearing headphones, while 65.9% did not.

Measures

Recordings

The study consisted of evaluating various versions of audio instructional recordings prepared for the study. The recordings were self-constructed specifically for the purposes of this study in the participants' native language (Polish language). Two native speakers—a woman and a man (both in their 20 s)—were recorded. The audio files were created using a Shure SM-58 dynamic microphone, a pop filter, and appropriate room soundproofing. The Shure SM-58 is a dynamic microphone designed for recording vocals, and

Table 1 Characteristics of the study participants

Variable	*n*	%
Age		
18–26	72	52.2%
27–39	16	11.6%
40–60	44	31.9%
Over 60	6	4.3%
Gender		
Female	87	63%
Male	48	34.8%
Other	3	2.2%
Ethnicity		
Caucasian	138	100%
Practicing mindfulness		
Yes	45	32.6%
No	56	40.6%
Not anymore	7	5.1%
I'm going to	30	21.7%

the pop filter reduces distortions in the recording caused by plosive sounds. Additionally, appropriate room soundproofing was utilized during the recording process. The recorded traces were de-noised and processed using Reaper software, which is a digital audio workstation used for recording, editing, and producing audio files. During processing, breathing noises, saliva sounds, and other artifacts were removed to produce the most neutral sound possible for the individual messages without distracting the listener (except the fourth segment, where the recordings differed in the number of artifacts present). In such cases, the recordings were left in their original form. For the third recording version, all possible artifacts and non-verbal noises were removed.

All recordings underwent equalization and gentle compression to improve their listening quality, using the VST plugin Izotope Nectar 3. Equalization is the process of adjusting the volume of certain frequency bands in an audio signal to achieve a specific effect. In this case, equalization was used to adjust the characteristics of the recording to a human voice. The recordings also underwent compression, which involves adjusting the volume of different parts of an audio signal, thus reducing its dynamic range. Izotope Nectar 3 is a multifunctional audio processing plug-in equipped with an equalizer, compressor, and other features, and is used in digital audio workstations.

The audio material included approximately 20-s excerpts from the introductory instructional segment for practice techniques such as mindful breathing and body scan (see Supplementary Information). The recording durations were limited to excerpt length (rather than the entire recording) to ensure adequate time to elicit responses while reducing the risk of study participant fatigue and dropout. Here is a sample transcript of the recording content (except for modifications due to manipulation):

> Take a comfortable seat where you can relax freely. Allow yourself a moment of peace during which no one will disturb you. Lie down comfortably so that you have support under your head. Close your eyes and begin to observe your breathing.

Rater 1, who has a Ph.D. in psychology and is a practitioner using mindfulness techniques in her practice, found that each passage met the conditions for a recording from a mindfulness training manual.

The study consisted of 24 recordings, with 12 featuring male voices and 12 featuring female voices. It was divided into four segments, each containing six voice recordings, with three variants of the recording in both female- and male-voiced versions (links to the recordings can be found in the Online Resource 1).

Grammatical Form The first segment comprised recordings with varying forms. The first recording featured a first-person message in the imperative mood, "Take a comfortable place where you can relax freely…" The second recording presented a message in the proposing form with words that softened the imperative mood, such as "If you want," "you can," and "if you feel the need." The third recording was in the imperative mood but in the first-person plural, "Let's take a comfortable place where we can relax freely…".

Pace The second segment included three recording variants with differing speech rates and pause lengths. The first recording was delivered at a fast rate with short pauses between sentences, requiring 25% less time to finish a phrase than at a moderate pace, including adjustment of the length of the pauses. The second variant was spoken at a moderate pace with medium-length pauses between sentences. The third variant was characterized by a slow speech rate with long between-sentence pauses, requiring 25% more time to finish a phrase than at a moderate pace, including adjustment of the length of the pauses.

Whispering The third segment included recordings with varying levels of whispering. The first recording was neutral, with no whispering; the second was partially whispered, with vocal cords tensed up to 25% of the maximum achieved by the speaker's fully whispered voice; and the third was primarily whispered, with vocal cords tensed up to 75% of the maximum achieved by the speaker's fully whispered voice.

Artifacts In the fourth and final segment, the recordings differed in the number of artifacts present. The first recording had many artifacts, with all of them left in the recording at the same volume as the speech. The second recording had a moderate amount of artifacts, with 25% of them removed and the volume of the remaining ones reduced by half. The third recording had no artifacts, making it analogous to a radio recording.

Rater 2, who is a speech acoustician and a university lecturer in the Department of Acoustics, evaluated the recordings carefully and precisely to ensure they had the necessary intensity levels of the specified properties.

The participants evaluated each recording by answering the question, "How do you rate this recording?" The ratings were given using a 5-point Likert scale, scored as follows: 1, *Dislike it very much*; 2, *Do not like it*; 3, *Neither like nor dislike it*; 4, *Like it*; 5, *Like it very much*. Participants were encouraged to answer an open-ended question: "Here you can leave comments on the above recordings—were you annoyed by any features of the messages? Or did you like something? Write your observations."

Data Analyses

Because of the hierarchical structure of our data (i.e., ratings nested participants), we used multilevel modelling to estimate hypothesized effects. We conducted four separate analyses for each of the following independent variables: form of the message (imperative vs plural vs propositional), pace of speech (slow vs medium vs fast), whispering (partial vs neutral vs full), and artifacts (no vs some vs many).

Each of the analysis was conducted in three steps. First, we included into the model only main effects of the manipulated variable (for each of the analyses it was subsequently form of the message, pace of speech, whispering, and artifacts), the instructor's gender and the participant's gender. At the second step, we added the effects of simple interactions between the predictors, and at the final step we added the complex interaction between three independent variables. At each of the steps, we controlled participant's meditation experience and whether a participant used headphones during the task. To control for rating variance due to individual differences and for variance of the instructor's gender effect due to individual differences, we allowed two parameters, i.e., the intercept as well as the instructor's gender effect to vary across participants. Next, we compared the models estimated at each of the analysis steps using χ^2 difference test. For the final description, we selected the model that significantly increased the explained variance of the dependent variable comparing the simpler model. All post-hoc tests were performed with Tukey's adjustment.

Results

For each of the manipulated variables (form of the message, pace of speech, whispering, and artifacts), we performed a separate analysis. Result of the four hierarchical multilevel regressions are presented in Table 2. Below we describe the results of these analyses in detail.

Form of the Message

Adding interactions to the model—either simple or complex—did not increase significantly explained variance of recording's rating ($\Delta\chi^2(5)=9.84$, $p=0.08$; and $\Delta\chi^2(2)=1.63$, $p=0.44$, respectively). Therefore, for further analysis; we selected the model with only the main effects included. The main effect of message form was significant: $F_{(2,132)}=10.51$, $p<0.001$ (Fig. 1). Post hoc tests with Tukey's adjustment revealed that instructions spoken in the plural form ($M=2.63$) were rated significantly lower compared to the imperative form ($M=3.08$; $t(133)=4.39$, $p<0.001$) and the propositional form ($M=2.97$; $t(133)=3.75$, $p<0.001$) that did not differ significantly from each other ($t(133)=1.21$, $p=0.45$). Whether participants used headphones or speakers or had mindfulness practice did not predict rating of recording significantly.

Pace of Speech

The model with simple interactions between pairs of independent variables was better fitted to data than the model including only main effects ($\Delta\chi^2(5)=35.58$, $p<0.001$). However, adding the complex interaction between all independent variables did not increase the explained variance of recording's rating ($\Delta\chi^2(2)=3.37$, $p=0.19$). Therefore, we present results of the second step of analysis. Both instructor's gender and pace of speech significantly influenced the rating of the recording: $F_{(1,334)}=19.37$, $p<0.001$ and $F_{(2,238)}=25.41$, $p<0.001$, respectively.

Furthermore, we observed a significant interaction effect between instructor's gender and pace of speech: $F_{(2,266)}=18.39$, $p<0.001$ (Fig. 2). Simple effect analyses revealed that instruction presented by a female instructor was significantly higher rated when they were spoken in a medium pace ($M=3.28$) compared to fast pace ($M=2.74$; $t(310)=5.43$, $p<0.001$) and slow pace ($M=2.53$; $t(266)=6.80$, $p<0.001$) that did not differ significantly from each other ($t(251)=1.83$, $p=0.16$). However, when the instructor's voice was male, instructions spoken with fast or medium pace were evaluated similarly ($M=3.26$, and $M=3.06$ respectively; $t(310)=1.98$, $p=0.12$) and both differed from recordings with slow pace ($M=2.46$; $t(251)=6.91$, $p<0.001$, and $t(266)=5.45$, $p<0.001$, respectively).

When the pace was fast, the instruction was rated higher if spoken with male than female voice ($t(373)=5.16$, $p<0.001$). The opposite effect was observed, however, if the pace was medium: instruction spoken with a female voice was rated higher than spoken with a male voice ($t(373)=2.19$, $p=0.029$). No difference between the female and male voices was observed in the instructions recorded with a slow pace ($t(373)=0.71$, $p=0.48$). No effects of headphones and mindfulness practice were observed.

Whispering

The model did not improve when simple or complex interactions were included ($\Delta\chi^2(5)=10.49$, $p=0.06$ and $\Delta\chi^2(2)=2.28$, $p=0.32$, respectively). Therefore, for further analysis, we selected the model with only the main effects included. The main effect of whispering was significant: $F_{(2,534)}=20.88$, $p<0.001$ (Fig. 3). Post hoc tests with Tukey's adjustment revealed that instructions spoken with full whispering ($M=2.57$) were rated significantly lower compared to the partial whispering ($M=2.99$; $t(534)=4.74$, $p<0.001$) and the neutral instructions ($M=3.12$;

Table 2 Hierarchical multilevel models for recording ratings

Variable	MV = Form				MV = Pace				MV = Whisper				MV = Artifacts			
	F (df)	df	R^2	ΔR^2	*F* (df)	df	R^2	ΔR^2	*F* (df)	df	R^2	ΔR^2	*F* (df)	df	R^2	ΔR^2
Step 1			0.58	0.58			0.51	0.51			0.30	0.30			0.34	0.34
Intercept	110.77	1.14			112.39	1.14			70.13	1.14			100.36	1.14		
Headphones	< 0.01	1.13			0.36	1.13			0.79	1.13			0.05	1.13		
Practice	0.54	3.13			0.14	3.13			0.57	3.13			0.61	3.13		
PS	0.01	1.13			1.22	1.13			0.97	1.13			0.16	1.13		
LS	3.08†	1.13			1.01	1.13			22.25***	1.13			0.92	1.13		
MV	10.51***	2.13			33.00***	2.13			20.88***	2.53			64.34	2.534***		
Step 2			0.59	0.01†			0.56	0.05***			0.31	0.01†			0.40	0.06***
Intercept	104.50	1.16			94.13	1.15			70.51	1.16			115.20	1.16		
Headphones	< 0.01	1.13			0.36	1.13			0.79	1.13			0.05	1.13		
Practice	0.54	3.13			0.14	3.13			0.57	3.13			0.61	3.13		
PS	0.03	1.13			1.26	1.13			0.87	1.29			0.12	1.25		
LS	1.71	1.33			19.37***	1.33			13.74***	1.46			36.41	1.44***		
MV	11.83***	2.21			25.41***	2.24			4.65**	2.53			9.84	2.53***		
PS × LS	0.26	1.13			0.31	1.13			0.11	1.13			2.55	1.13		
PS × MV	1.52	2.13			0.33	2.13			0.07	2.53			5.51	2.53**		
LS × MV	3.37**	2.27			18.39***	2.27			5.11**	2.53			23.20	2.53***		
Step 3			0.59	0.00			0.57	0.01			0.31	0.00			0.40	0.00
Intercept	101.32	1.16			92.12	1.15			67.69	1.16			117.21	1.16		
Headphones	< 0.01	1.13			0.36	1.13			0.79	1.13			0.05	1.13		
Practice	0.54	3.13			0.14	3.13			0.57	3.13			0.61	3.13		
PS	0.09	1.18			0.59	1.19			1.44	1.43			0.13	1.37		
LS	2.91†	1.34			17.96***	1.38			8.38**	1.56			35.99***	1.55		
MV	9.73***	2.24			18.66***	2.29			5.24**	2.53			5.97**	2.53		
PS × LS	0.33	1.38			0.02	1.38			0.68	1.56			4.42*	1.55		
PS × MV	1.19	2.24			0.07	2.29			0.44	2.53			1.29	2.53		
LS × MV	2.42†	2.26			13.53***	2.26			4.72**	2.53			20.04***	2.53		
PS × LS × MV	0.80	2.26			1.67	2.26			1.13	2.53			1.02	2.53		

Practice, mindfulness practice; *PS*, participant's sex; *LS*, lector's sex; *MV*, manipulated variable

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; "†" a trend, $p < 0.01$

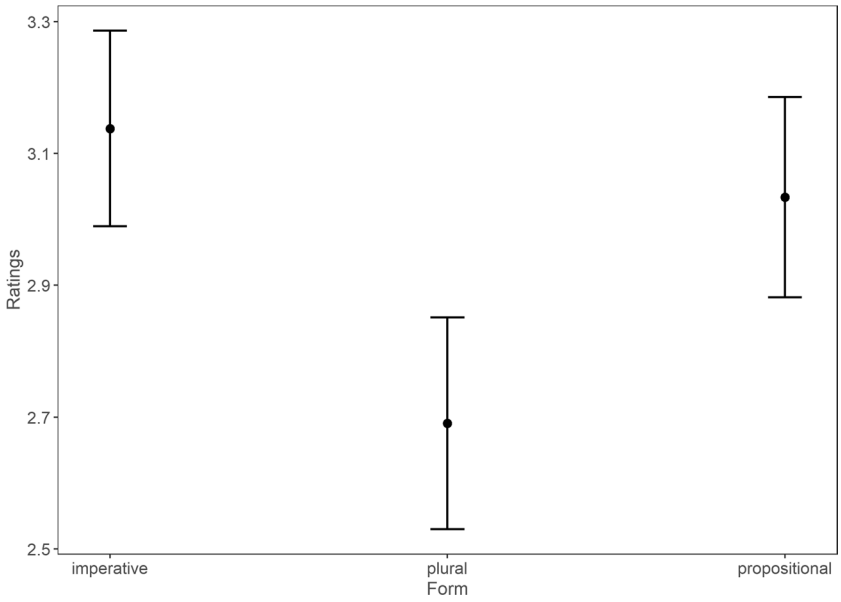

Fig. 1 Recording rating as a function of the form of message

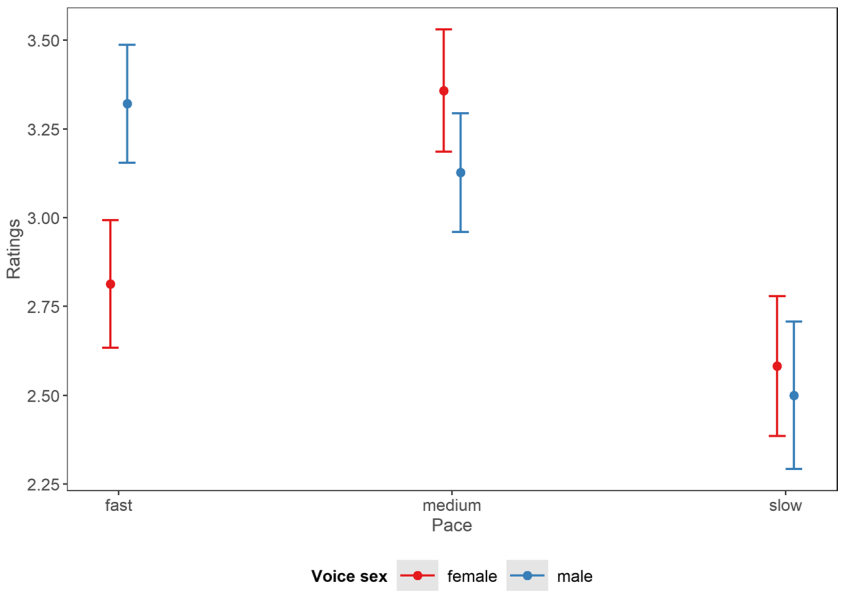

Fig. 2 Recording rating as a function of the interaction between pace of speech and voice gender

$t(534) = 6.17$, $p < 0.001$) that did not differ significantly from each other ($t(534) = 1.43$, $p = 0.33$).

The main effect of instructor's gender was also observed: $F_{(1,133)} = 22.25$, $p < 0.001$. Recordings with a female voice ($M = 3.08$) were rated higher compared to male voice ($M = 2.71$). Whether participants used headphones or speakers or had mindfulness practice did not predict rating of recording significantly.

Artifacts

The model with simple interactions between pairs of independent variables was better fitted to data than the model including only main effects ($\Delta\chi^2(5) = 57.69$, $p < 0.001$). However, adding the complex interaction between all independent variables did not increase the explained variance of recording's rating ($\Delta\chi^2(2) = 2.06$, $p = 0.36$). Therefore, we present results of the second step of analysis. Both instructor's gender and artifacts influenced rating of recording significantly: $F_{(1,439)} = 36.41$, $p < 0.001$ and $F_{(2,530)} = 9.84$, $p < 0.001$, respectively.

Likewise, we observed a significant interaction effect between instructor's gender and artifacts: $F_{(2,530)} = 23.20$, $p < 0.001$ (Fig. 4). Simple effect analyses revealed that the instruction presented by a female instructor was rated highest if it included some artifacts ($M = 3.19$) compared to many

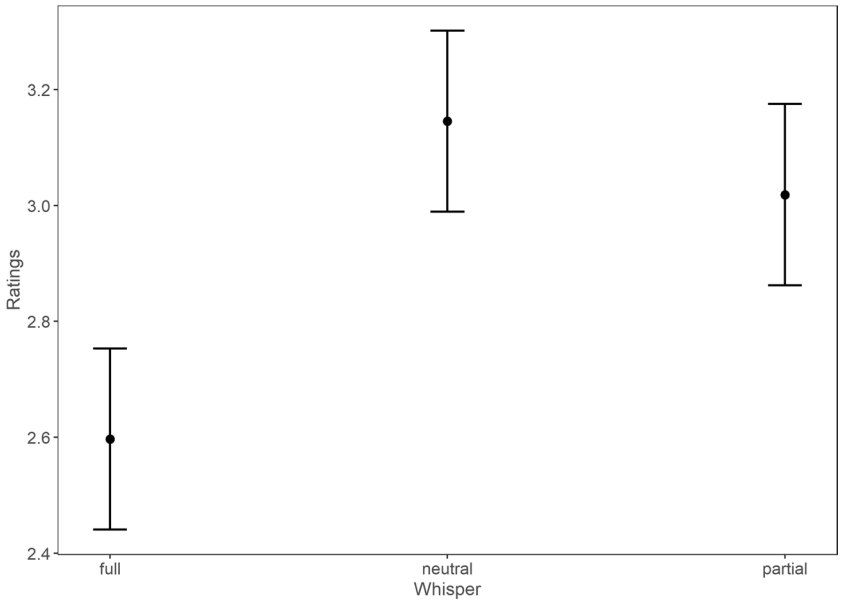

Fig. 3 Recording rating as a function of the level of whispering

artifacts ($M = 2.80$; $t(530) = 3.69$, $p < 0.001$) and no artifacts ($M = 2.95$; $t(530) = 2.31$, $p < 0.06$) that did not differ significantly from each other ($t(530) = 1.38$, $p = 0.36$). However, when the instructor's voice was male, instructions including no or some artifacts were rated similarly ($M = 3.27$, and $M = 3.36$ respectively; $t(530) = 0.87$, $p = 0.66$) and both differed from recordings with many artifacts ($M = 2.22$; $t(530) = 10.14$, $p < 0.001$, and $t(530) = 11.02$, $p < 0.001$, respectively).

When instructions included many artifacts, they were rated higher if spoken with a female than a male voice ($t(528) = 5.49$, $p < 0.001$). The opposite effect was observed, however, if instructions included no artifacts: recordings of male voice were rated higher than of female voice ($t(528) = 3.02$, $p = 0.003$). No difference between the female and male voices was observed in the instructions including some artifacts ($t(528) = 1.63$, $p = 0.10$).

The interaction effect between participants' gender and artifacts was also significant: $F_{(2,530)} = 5.51$, $p < 0.001$ (Fig. 5). For male participants, all conditions differed significantly from each other, i.e., instructions with many artifacts ($M = 2.59$) were rated lower than those with no artifacts ($M = 2.94$; $t(530) = 2.9$, $p = 0.011$) and those with some artifacts ($M = 3.26$; $t(530) = 5.54$, $p < 0.001$); and instructions with no artifacts were rated lower than instructions with some artifacts ($t(530) = 2.64$, $p = 0.023$). For female

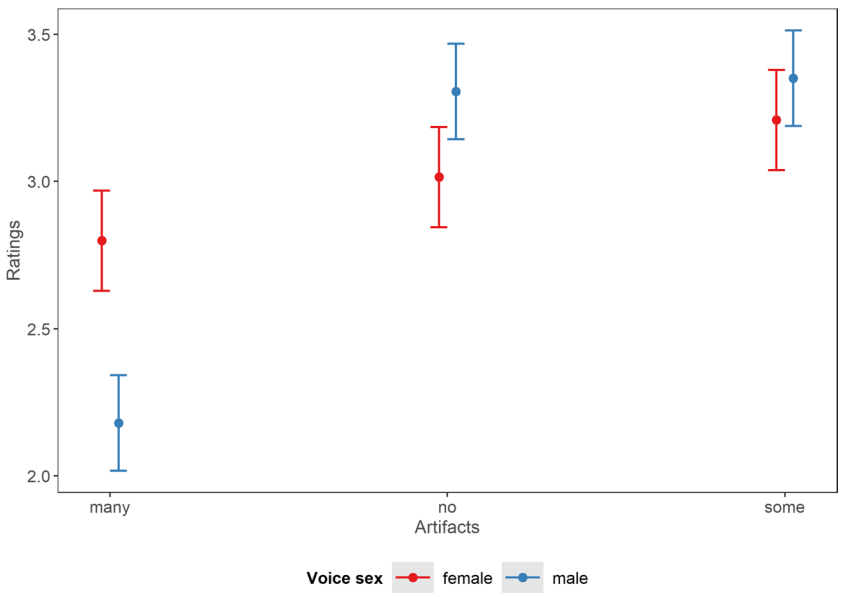

Fig. 4 Recording rating as a function of the interaction between number of artifacts and voice gender

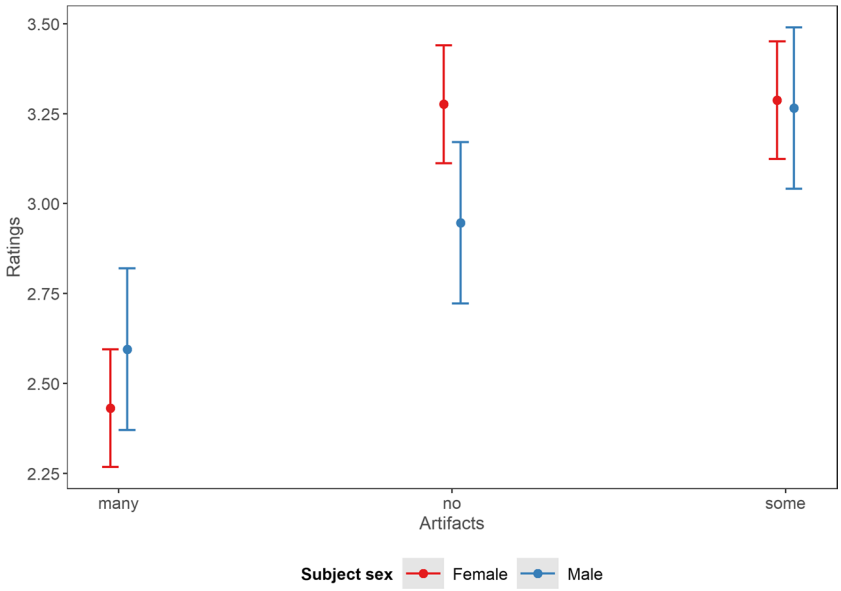

Fig. 5 Recording rating as a function of the interaction between number of artifacts and participant's gender

participants, there was no difference between instructions with no and some artifacts ($M = 3.27$ and $M = 3.29$, respectively; $t(530) = 0.129$, $p = 0.99$); however, ratings in those conditions were significantly higher compared to the condition with many artifacts ($M = 2.43$; $t(530) = 9.50$, $p < 0.001$ and $t(530) = 9.63$, $p < 0.001$, respectively). Ratings of male and female participants differed only in the condition with no artifacts ($t(296) = 2.31$, $p = 0.02$). No effects of headphones and mindfulness practice were observed.

Responses to Open-Ended Questions

Open-ended feedback was collected from participants after each segment in the study (Segment I, 61; Segment II, 49; Segment III, 43; Segment IV, 44). The initial responses were organized into four categories, corresponding to the study's four segments. To identify thematic categories, we conducted open coding, looking for the most frequently discussed topics within each segment's responses. Consequently, participant responses were divided among categories associated with the recordings' most commonly assessed properties, as well as an "other" category for individual ratings of specific properties. A comprehensive content analysis was carried out within each category to differentiate between positive and negative evaluations of the recordings. For instance, in the category focused on communication pace, subcategories were established, such as negative evaluations of the slowest recordings and negative evaluations of lengthy pauses between words. Additional individual responses addressing various aspects of the recordings were incorporated into the "other" subcategory.

Content analysis revealed that a majority of the statements possessed a distinct emotional undertone, primarily of a negative nature. This was evident from the participants' use of capital letters, emoticons, and exclamation points to underscore their sentiments. All quotes have retained the original spelling from the participants' statements.

In the first segment, 43% of the comments were negative remarks about the use of "we" (first-person plural). According to respondents, this form was decidedly unappealing and introduced chaos and misinformation into the perception of the message. The following are sample statements reflecting criticism of this recording:

I was irritated by the messages spoken 'to us.'

When they speak in the plural, it is creepy.

In the second segment, recordings with a slow pace were the most heavily criticized (43%). The majority described the recordings' soporific qualities or their feelings of impatience. For example:

Too slow delivery of the message (sleepy). The trainee will not make it to the end because he will fall asleep.
The slow ones, more than relaxing, make a person feel like throwing the phone out the window.

Some people (14%) also did not like excessively long pauses between sentences:

I do not like pauses between messages.
When we want to give the listener enough of that time
i t i s g e t t i n g a w f u l l y s n a z z y y y

In the third segment, almost half of the respondents (47%) viewed whispering unfavorably. They stated that whispering seemed to them to be a communication that crossed the boundary of intimacy and was somewhat erotic, introducing discomfort and a lack of naturalness, such as:

> The male voice was too 'emotional' for me, maybe even erotic? It made me feel uncomfortable, tense.
> Is it still a practice, or already a gender phone?

The participants also stated that whispering made them sleepy. Therefore, it would be challenging to practice with such a voice, for example:

> Silent messages are sleepy, and I don't think putting people to sleep is the point.

In segment 4, more than half of the responses (64%) described adverse reactions to recordings with clearly audible artifacts. They were described as annoying, distracting, and strange, for example:

> Disgusting munching yuck.
> CIAMK CIAMK CIAMK CIAMK (the sounds of smacking lips)
> The sounds of swallowing saliva and breathing seemed irritating. I would not be able to concentrate with them.

Some respondents indicated that such recordings could be uncomfortable for people with misophonia. They suggested introducing this information in future studies.

Respondents also described additional characteristics of the voices that influenced the perception of the recordings. Some comments (16%) concerned vocal properties, such as timbre, intonation, diction, or vigor of the utterance, for example:

> The message is too 'harsh' in reception.
> The voices were too loud. When practicing mindfulness, a calm, quiet, composed voice is important.

Several individuals expressed their preference for a lower voice timbre. Additionally, the absence of a strong accent and/or exaggerated diction was also considered attractive. Some respondents highlighted that a "positive attitude in the voice" had a favorable impact on the recording's appeal:

> I definitely prefer recordings where the way of speaking reveals the calmness and positive attitude of the speaker.
> Nevertheless, some comments on a particular feature were positive, for example, approving the slow pace. Some opinions approved the propositional recording form. One respondent pointed out that this type of recording is often preferred by stressed people. The propositional recording form, which offers suggestions rather than direct commands, builds a sense of internal control. This approach is seen as more helpful compared to the "imperative mood"—a directive or command-like style of instruction—which was described by a number of respondents as aggressive. Some people explained in detail which message features they found most desirable. They also described the features that they felt were lacking (for example, calm background music):
> There was nothing to encourage relaxation in any of the recordings. It would have to be a completely different timbre of voice and end with light relaxing music.
> There was no quiet background music

Discussion

The present study aimed to analyze the influence of selected characteristics of recorded mindfulness practice instructions on recipients' evaluation. The study particularly focused on elements that spur negative reactions, which may hinder mindfulness practice. The results showed that recipients' gender did not affect their preference for male- or female-voiced recordings. However, the properties of recordings such as form, pace, whispering, and artifacts affected participant evaluation. We identified message elements that listeners liked least. These include using plural pronouns, slow pace, primarily whispering, and a high number of distracting sounds or "artifacts." Moreover, some differences in evaluations were noted between recordings with female voices and those with male voices. In addition, more detailed qualitative data were collected on the participants' subjective reactions, including their reflections and emotions when exposed to recordings with different characteristics. At the same time, we identified elements of the preferred versions.

Overall, the findings suggest that the form of the message, the pace of speech, whispering, artifacts, and the instructor's gender may have a significant impact on the perceived quality of recorded meditation instructions. We found no significant effect of age and meditation experience on preference for specific message properties. Whether participants used headphones was also irrelevant to their responses, indicating that neither technical considerations nor personal experience has an impact on the listeners' reaction.

Contrary to our initial assumptions, the study's results indicated that participant gender did not have a significant impact on ratings in general. This finding is noteworthy because it challenges previous studies that suggested a preference for the opposite gender's voice (Xu et al., 2013). Furthermore, it does not align with data showing a general trend of higher ratings for female voices in command situations (Parson et al., 2013). These findings could be due to the unique context of mindfulness instruction, which might create distinct experiences compared to those in prior research. However, a detailed analysis of the ratings for each condition uncovered differing tendencies toward pacing, whispering, and the number of artifacts in female and male voice recordings. These observations support earlier claims that sociolinguistic patterns can vary across different contexts and that

expectations about language use may change in response to situational factors (Fitzpatrick et al., 1995).

In line with our expectations, the unique features of the mindfulness recordings in each section of the survey did influence the varied responses from listeners. First, the results indicated that instructions using plural forms received significantly lower ratings compared to those using imperative and propositional forms. This finding aligns with previous studies suggesting that feelings of connection or intimacy with the instructor are not preferred in mindfulness training (Slatcher et al., 2008). Interestingly, no significant difference was observed between imperative and propositional forms, implying they are equally favored. This outcome challenges prior literature, which reported a stronger aversion to messages conveyed in imperative forms (Wicklund, 1974). This result could be attributed to the fact that the imperative mood is commonly used in mindfulness and relaxation techniques and many individuals have grown accustomed to it (Rothschild, 2017). Furthermore, no differences in preferences were found between male and female versions of the recordings. This suggests that the instructional form elicits similar reactions in both genders, irrespective of the speaker's gender. Our findings challenge established research suggesting that listeners expect specific language patterns; for instance, girls are perceived to often use the imperative mood in the first person plural when making suggestions, while boys are more likely to issue orders to one another (Tannen, 1986). Once again, the context of mindfulness practice may alter reactions.

Second, the study's findings highlight the importance of speech speed in how participants rate recorded instructions. Specifically, instructions delivered at a moderate pace received the highest ratings, irrespective of the speaker's gender. However, for female speakers, instructions spoken at a slow pace were rated lower than those at a moderate pace, but similar to those at a fast pace. This indicates that neither slow nor fast pacing is appealing to listeners when the instructor is a woman. Conversely, for male speakers, instructions spoken at a slow pace were rated lower than those at a moderate or fast pace, but there was no significant difference between ratings for instructions spoken at a fast or moderate pace. This suggests that when listening to a man, people specifically dislike slow pacing, while fast and moderate pacing may be preferred. Additionally, when the pace was fast, the instruction was rated higher if spoken with male than female voice, and the opposite effect was observed when the pace was medium. In partial alignment with our hypothesis, a slow pace was less preferred, but this was only the case when the lecturer's voice was male. When the voice was female, the slow pace did not significantly differ from the fast pace, which also received low preference ratings. This outcome is mainly consistent with reports linking aversion to slow messages to difficulty maintaining attention (Boone et al., 2019). It also supports the data on the tendency to direct attention to formal message properties in situations where conditions are perceived as too slow, such as slow speaking or reading rates (Belin et al., 2011).

When the pace of speech was fast, instructions spoken with a male voice received higher ratings, while the opposite effect was observed when the pace was moderate. These findings partially support our hypothesis, which suggests that a slow pace would be less preferred. However, this was only observed when the speaker's voice was male, as there was no significant difference between a fast or slow pace when the speaker was female. These results align with previous reports, which suggest that people tend to have difficulty maintaining attention when messages are delivered slowly (Boone et al., 2019). They also support data on the tendency to focus on more formal message properties in such conditions (Belin et al., 2011).

The variations in participants' preferences for pacing based on voice gender may be attributed to gender-specific language styles and societal expectations for communication between women and men (Fitzpatrick et al., 1995; Murachver & Janssen, 2007). Distinct differences in the way women and men communicate can be observed at every level of language and speech organization (Carli, 1990; Kloch, 2000). Once individuals are identified as members of certain social categories, they are expected to behave according to the norms associated with those categories. A fast pace of speech appears to align more closely with the male-typed directive approaches to conversation, which are often competitive and argumentative, and was therefore better received in male recordings than in female ones (Murachver & Janssen, 2007). This finding is in line with a study by Janssen (2004), which reported that ratings were consistently lower for female-labeled speakers who exhibited male-typed conversation styles. Women who speak like men tend to be judged more negatively than men who use the same conversational style.

Third, consistent with our assumptions, the results demonstrate that whispering has a significant impact on the ratings of the recordings, with full whispering receiving lower ratings compared to partial whispering and neutral instructions. This finding implies that using full whispering in instructional recordings may adversely affect their perceived quality, while natural speech and partial whisperings are equally favored. This trend could stem from the inappropriateness of the chosen communication style in an instructional context and an overly intense sense of intimacy that may trigger anxiety (Janik McErlean & Banissy, 2018).

Moreover, the study identified a significant main effect of the speaker's gender on the ratings of the recordings, with female voices scoring higher compared to male voices. This outcome aligns with previous research that found female voices are generally considered more pleasant and soothing

than male voices (Dolliver, 2010). Although earlier studies suggested that a female whisper was perceived as erotic (Hughes et al., 2014), our research indicated that the male version was rated lower. This discrepancy may be attributed to content addressing bodily sensations, which could evoke stronger negative associations due to concerns of boundary transgressions by men.

Finally, the study discovered that artifacts significantly influenced the ratings of recordings. As anticipated, participants showed little preference for recordings with numerous artifacts, consistent with research on reactions to such sounds and the increasingly recognized phenomenon of misophonia (Janik McErlean & Banissy, 2018). Additionally, a significant interaction effect was observed between the speaker's gender and artifacts. Instructions delivered by a female speaker received the lowest ratings when they included many artifacts compared to some artifacts. However, there was no significant difference in ratings between recordings with many artifacts and those completely free of artifacts (removed in post-production). This result suggests that the removal procedure might also elicit negative reactions when the instruction is given by a woman. Recordings with a typical number of artifacts by female speakers are preferred the most.

Conversely, when the speaker's voice was male, recordings with numerous artifacts were also less favored, but instructions with no or some artifacts received similarly high ratings. Thus, the removal procedure appears unnecessary to satisfy practitioners when instructions are given by a man (Cavanna & Seri, 2015). Instructions with many artifacts were rated lower when spoken by a male voice compared to a female voice. However, the opposite effect was observed for instructions with no artifacts: participants preferred male voices over female voices. Ratings of instructions with some artifacts did not differ between female and male voices.

There was also a significant interaction effect between artifacts and the participants' gender. Specifically, male participants rated instructions with numerous artifacts lower than those with some or no artifacts. Additionally, instructions without artifacts were marginally less preferred by men compared to the natural version. In contrast, female participants did not show any significant difference in ratings between instructions with no and some artifacts. However, their ratings in these conditions were significantly higher than those in the condition with many artifacts. The only condition in which male and female participants' ratings differed was in the absence of artifacts. This suggests that both genders perceive and react to artifacts differently, with women being more sensitive or more affected by the presence of numerous artifacts. This finding aligns with research demonstrating that women manifest significantly higher severity of misophonia symptoms than men, and female gender predicts higher misophonic distress (Smith et al., 2022; Stalias-Mantzikos et al., 2023).

Results of this study obtained through responses to open-ended questions and exploratory analysis provided additional complementary qualitative data. Precise and detailed reflections suggest high awareness of message qualities that are least and most desirable. They also showed a wide range of reactions to undesirable features of the recordings. These features may hinder engagement in mindfulness training. The content of the responses was mostly consistent with the statistical analysis results. Most of them reflected trends evident in the recording evaluations. Respondents most frequently pointed out the undesirable use of first-person plural, the recording being too slow-paced, and the unnecessarily whispered forms of the messages, as well as the excessive, distracting number of auditory artifacts in the recordings. Nonetheless, we discovered a few differing opinions from the participants, suggesting that some individual variations should be considered in evaluating possible external obstacles in instructional features of mindfulness training. Moreover, qualitative analysis of the open-ended question responses identified additional properties not considered in the hypotheses. These features may also hinder the reception of messages given during mindfulness training. Participants disliked excessive vocal energy and unfavorable diction, preferring a calm tone and soft utterances. They also disliked overly expressive intonation. These findings should be further investigated.

Overall, the results of the study provide insights into the impact of specific vocal characteristics on the perceived quality of instructional recordings. The study revealed the general preferences of potential and current mindfulness training participants regarding selected aspects of voice instruction. The findings suggest that the form of the message, the pace of speech, whispering, and artifacts are important factors to consider when delivering instructions, as they can significantly affect how they are perceived by the audience. The gender of the instructor may also play a role in how instructions are received.

This study focused on the external challenges of practicing mindfulness and aimed to support individuals in their mindfulness practice to improve their overall functioning. Our findings identified specific voice recording attributes that might deter participants from engaging or prompt them to discontinue the practice, even when they recognize the potential benefits of mindfulness. Mindfulness practice involves developing an accepting attitude towards all experiences, even difficult ones (Kabat-Zinn, 2003; Williams et al., 2007). However, exposure to challenging experiences can make practicing mindfulness more difficult, especially for beginners or those with heightened sensitivity or cognitive-emotional difficulties.

While mindfulness aims to increase the ability to accept challenging experiences, negative experiences can hinder commitment to training. Studies have shown that people with low levels of commitment tend to abandon their attempts after facing initial challenges (Germer et al., 2005). This problem is similar to physical training under challenging conditions, where difficult circumstances can both enhance acquired skills and discourage continued practice. It is crucial for practitioners to find a balance that promotes development without exceeding their capabilities, which may lead to quitting. Our study can help create conditions conducive to developing mindfulness skills by reducing risk factors associated with quitting, thus supporting individuals in their mindfulness journey.

At the same time, the study discovered the most desirable aspects of the recordings. The findings can be used to inform the design of instructional materials and to train narrators to be more effective communicators. Information on hindrances and preferred factors are essential for the practical design of mindfulness training methods. The present study's results can support the creation of better quality digital forms of mindfulness practice, with attention to the needs and preferences of practitioners. They can help create better, more tailored recordings available in apps or platforms for individual preferences while reaching the largest audience. Such modern solutions increase mindfulness training's accessibility. These solutions will be valuable for people who prefer to exercise using a computer or phone at a convenient time and place (Mrazek et al., 2019).

The results may also contribute to developing more personalized practices. Recordings features could be selected or self-modified according to individual preferences. These might include voice gender, form, or amount of whispering (some changes, such as pace or adding background music, are already possible in applications). This option could be of particular interest to people with specific preferences beyond the most common response tendencies (e.g., those with misophonia, attention deficits, psychological trauma, and others). The results support the creation of diverse recordings so that people with special preferences have practice opportunities (especially if they find it challenging to carry out mindfulness training without voiceover assistance). Thus, these study findings may facilitate mindfulness practice for many people while increasing the choice, attractiveness, and customization of mindfulness recordings available on courses, platforms, and applications. Such improvements may lead to better training outcomes and increased practice effectiveness (Gál et al., 2021; Mrazek et al., 2019).

In addition, mindfulness practice recordings are instruments used in scientific research on mindfulness (Altschuler et al., 2012; Walsh et al., 2019). Awareness of negatively perceived recordings' features may improve these tools and the accuracy of research results. Thus, our study results can also serve to develop and improve the quality of instruction, streamline the research process, and increase the chances of success.

Finally, the study can also serve to inspire trainers. The data revealed the significant role that specific instruction characteristics play and showed the reactions in the listeners/receivers of the practice. During in-person practice, additional factors, such as the instructor's appearance, the instructor-participant relationship, and the presence and the reactions of other participants, may influence the participants' evaluations and reactions. Nevertheless, taking into account the data obtained may promote more conscious practice and care toward the elements of the instruction that are crucial from the perspective of the participant's experience.

Limitations and Future Directions

The current study has some limitations that must be considered for future research. Firstly, the survey was conducted remotely and there were variations in the survey conditions. Some factors could not be controlled, such as noise, time of day, or the presence of other people in the room during the administration of the survey. Thus, the participants' responses may have been affected. It would be worthwhile to consider standardizing the research conditions and comparing laboratory and remote study results in future research. Despite this, it is important to note that the varied conditions were consistent with the realistic situations in which people practice mindfulness with the use of recordings (e.g., at home, in the hospital, at work, on the train, on a walk, at different times of the day); therefore, the ecological relevance of the survey may be better using the current study procedure.

A second technical issue worth considering is the content and length of the recordings, which may have had an additional impact on the evaluation of the messages. In future research, it could be helpful to compare the effect of other content and longer excerpts of instruction on listeners' evaluations.

Thirdly, the results may have reflected the limited number of voices used (one female/one male) and their distinctive characteristics. Respondents may have had different preferences related to the specific voices used in the recordings. Some participants shared opinions about additional factors (e.g., the irritating timbre of voices). These vocal features can directly disturb the perception of the manipulated recording elements. Therefore, they can affect the content validity of the experiment. Future studies should introduce a more diverse selection of gendered voices and average the ratings to confirm this result.

Furthermore, while we based our selection of factors on the existing literature, it is important to note that there

may be other factors beyond the recording quality itself that can affect the perception of the meditation practice, such as accent, cultural context, or the subjectivity of listener bias. For example, accent can significantly influence speech perception, potentially leading to comprehension difficulties and negative emotions (Gluszek & Dovidio, 2010). Additionally, cultural differences, including language-related words and pronouns, may impact message interpretation (Matsumoto & Juang, 2016). Mindfulness meditation's roots in Eastern cultures could also affect the ease of focusing attention and tolerance for recording properties among participants from different backgrounds. The listener's subjective attitude, including expectations, attitudes toward the practice, and previous experiences, can influence meditation perception (Laurie & Blandford, 2016). So these findings are based on the specific sample of this study and may not generalize to other populations. Future research should examine the effects of these factors on the experiences of individuals from various practitioner groups.

Finally, the study was limited to analyzing recorded instructions. Future studies could compare participant evaluations of instructions delivered live by instructors at mindfulness trainings with instructions delivered via the recordings. Examining the impact of the instructor's presence, the ambiance of natural conditions, and the disposition or motivation of participants in real-time trainings might provide intriguing insights into factors shaping the perception of mindfulness guidance under varied circumstances.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12671-023-02264-8.

Acknowledgements We would like to thank Editage (http://www.editage.com) for English language editing.

Author Contribution Maja Stanko-Kaczmarek: conceptualization, project administration, methodology, formal analysis, writing–original draft, writing–review and editing, supervision. Tomasz Jankowski: formal analysis, writing–review and editing. Anna Koczyk: conceptualization, methodology, investigation,formal analysis, writing–original draft, writing–review and editing. Wadim Krzyzaniak: conceptualization, investigation, formal analysis, writing–original draft, writing–review and editing. Kalina Meder: conceptualization, formal analysis, writing–original draft, writing–review and editing. Kajetan Dulian: conceptualization, investigation, writing–review and editing.

All the authors approved the final version of the manuscript for submission.

Funding Not applicable.

Data Availability All data are available at the Open Science Framework (https://osf.io/r63v5/).

Declarations

Conflict of Interest The authors declare no competing interests.

Ethics Approval This study was approved by the Adam Mickiewicz University's Ethics Committee for Research Projects of the Faculty of Psychology and Cognitive Science.

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

Use of Artificial Intelligence AI was used for editing the manuscript to improve English language.

References

Altschuler, A., Rosenbaum, E., Gordon, P., Canales, S., & Avins, A. L. (2012). Audio recordings of mindfulness-based stress reduction training to improve cancer patients' mood and quality of life—a pilot feasibility study. *Supportive Care in Cancer, 20*(6), 1291–1297. https://doi.org/10.1007/s00520-011-1216-7

Andersen, J. (2014). Now you've got the shiveries: Affect, intimacy, and the ASMR whisper community. *Television & New Media, 16*(8), 683–700. https://doi.org/10.1177/1527476414556184

Bady, S. L. (1985). The voice as a curative factor in psychotherapy. *Psychoanalytic Review, 72*(3), 479–490.

Baminiwatta, A., & Solangaarachchi, I. (2021). Trends and developments in mindfulness research over 55 years: A bibliometric analysis of publications indexed in web of science. *Mindfulness, 12*(9), 2099–2116. https://doi.org/10.1007/s12671-021-01681-x

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Bhatia, S., Edwards, S., Stegner, J., & McCrickard, D. S. (2005). Do you hear voices? Exploring notification characteristics of voice. *Eleventh International Conference on Human-Computer Interaction*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=27d0b5c4047df0c397bf759e7d67f9b1159a0701

Blanck, P. D., Rosenthal, R., Vannicelli, M., & Lee, T. D. (1986). Therapists' tone of voice: Descriptive, psychometric, interactional, and competence analyses. *Journal of Social and Clinical Psychology, 4*(2), 154–178. https://doi.org/10.1521/jscp.1986.4.2.154

Boone, D., McFarlane, S., Berg, V. S., & Zraick, R. (2019). *The voice and voice therapy* (10th ed.). Pearson Education.

Britton, W. B., Lindahl, J. R., Cooper, D. J., Canby, N. K., & Palitsky, R. (2021). Defining and measuring meditation-related adverse effects in mindfulness-based programs. *Clinical Psychological Science, 9*(6), 1185–1204. https://doi.org/10.1177/2167702621996340

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of*

Personality and Social Psychology, 84(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822
Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology, 59*(5), 941. https://doi.org/10.1037/0022-3514.59.5.941
Cavanna, A. E., & Seri, S. (2015). Misophonia: Current perspectives. *Neuropsychiatric Disease and Treatment, 11*, 2117. https://doi.org/10.2147/ndt.s81438
Chittaro, L., & Vianello, A. (2016). Evaluation of a mobile mindfulness app distributed through on-line stores: A 4-week study. *International Journal of Human-Computer Studies, 86*, 63–80. https://doi.org/10.1016/j.ijhcs.2015.09.004
Ding, H., Hoffmann, R., & Jokisch, O. (2017). Prosodic correlates of voice preference in Mandarin Chinese and German: A cross-linguistic comparison. *Studientexte zur Sprachkommunikation: 28. Konferenz Elektronische Sprachsignalverarbeitung, 2017*, 83–90.
Dolliver, M. (2010). How people perceive gender in voiceovers. *Adweek, 51*(10), 18–18.
Economides, M., Martman, J., Bell, M. J., & Sanderson, B. (2018). Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: A randomised controlled trial. *Mindfulness, 9*(5), 1584–1593. https://doi.org/10.1007/s12671-018-0905-4
Farias, M., Maraldi, E., Wallenkampf, K. C., & Lucchetti, G. (2020). Adverse events in meditation practices and meditation-based therapies: A systematic review. *Acta Psychiatrica Scandinavica, 142*(5), 374–393. https://doi.org/10.1111/acps.13225
Fitzpatrick, M. A., Mulac, A., & Dindia, K. (1995). Gender-preferential language use in spouse and stranger interaction. *Journal of Language and Social Psychology, 14*(1–2), 18–39. https://doi.org/10.1177/0261927X951410
Gál, É., Ștefan, S., & Cristea, I. A. (2021). The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: A meta-analysis of randomised controlled trials. *Journal of Affective Disorders, 279*, 131–142. https://doi.org/10.1016/j.jad.2020.09.134
Germer, C. K., Siegel, R., & Fulton, P. (2005). *Mindfulness and psychotherapy*. Guilford Press.
Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review, 14*(2), 214–237. https://doi.org/10.1177/1088868309359288
Goldberg, S. B., Tucker, R. P., Greene, P. A., Simpson, T. L., Kearney, D. J., & Davidson, R. J. (2017). Is mindfulness research methodology improving over time? *A systematic review. PLoS ONE, 12*(10), e0187298. https://doi.org/10.1371/journal.pone.0187298
Hain, H. U., Jokisch, O., & Coelho, L. (2009). Multilingual voice analysis: Towards prosodic correlates of voice preference. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung* (pp. 215–221). TUDpress.
Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior, 38*, 107–127. https://doi.org/10.1007/s10919-013-0163-z
Janik McErlean, A. B., & Banissy, M. J. (2018). Increased misophonia in self-reported autonomous sensory meridian response. *PeerJ, 6*, e5351. https://doi.org/10.7717/peerj.5351
Janssen, A. L. (2004). Communication in context: How writers, speakers, readers, and listeners do gender. PhD thesis, University of Otago, Dunedin, New Zealand.
Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. Hyperion.
Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice, 10*(2), 144–156. https://doi.org/10.1093/clipsy.bpg016
Kabat-Zinn, J. (2017). Common obstacles to practice. *Mindfulness, 8*(6), 1713–1715. https://doi.org/10.1007/s12671-017-0820-0
Kabat-Zinn, J. (2013). *Full catastrophe living, revised edition: How to cope with stress, pain and illness using mindfulness meditation*. Hachette UK.
Kabat-Zinn, J. (2018). *Falling awake: How to practice mindfulness in everyday life*. Hachette UK.
Keng, S. L., Smoski, M. J., & Robins, C. J. (2011). Effects of mindfulness on psychological health: A review of empirical studies. *Clinical Psychology Review, 31*(6), 1041–1056. https://doi.org/10.1016/j.cpr.2011.04.006
Kloch, Z. (2000). Language and gender: Social and psychological determinants in communication. *Psychology of Language and Communication, 4*(2), 45–58.
Kumar, S., Tansley-Hancock, O., Sedley, W., Winston, J. S., Callaghan, M. F., Allen, M., Cope, T. E., Gander, P. E., Bamiou, D.-E., & Griffiths, T. D. (2017). The brain basis for misophonia. *Current Biology, 27*(4), 527–533. https://doi.org/10.1016/j.cub.2016.12.048
Laurie, J., & Blandford, A. (2016). Making time for mindfulness. *International Journal of Medical Informatics, 96*, 38–50. https://doi.org/10.1016/j.ijmedinf.2016.02.010
Loewy, J. (2004). Integrating music, language and the voice in music therapy. *Voices: A World Forum for Music Therapy, 4*(1). https://doi.org/10.15845/voices.v4i1.140
Loui, P., Bachorik, J. P., Li, H. C., & Schlaug, G. (2013). Effects of voice on emotional arousal. *Frontiers in Psychology, 4*, 675. https://doi.org/10.3389/fpsyg.2013.00675
Matsumoto, D., & Juang, L. (2016). *Culture and psychology*. Cengage Learning.
McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS ONE, 9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779
Mrazek, A. J., Mrazek, M. D., Cherolini, C. M., Cloughesy, J. N., Cynman, D. J., Gougis, L. J., Landry, A., Reese, J., & Schooler, J. W. (2019). The future of mindfulness training is digital, and the future is now. *Current Opinion in Psychology, 28*, 81–86. https://doi.org/10.1016/j.copsyc.2018.11.012
Murachver, T., & Janssen, A. (2007). Gender and communication in context. In W. Ann, M. W. Bernadette, & G. Cindy (Eds.), *Language, Discourse and Social Psychology* (pp. 185–205). Palgrave Macmillan. https://doi.org/10.1057/9780230206168_8
Murry, T. (1988). Vocal tract parameters associated with voice quality and preference. *Journal of Voice, 2*(2), 111–117. https://doi.org/10.1016/s0892-1997(88)80066-3
Parson, J., Braga, D., Tjalve, M., & Oh, J. (2013) Evaluating voice quality and speech synthesis using crowdsourcing. In: Habernal I., Matoušek V. (Eds.), *Text, speech, and dialogue. TSD 2013. Lecture Notes in Computer Science, 8082*. Springer. https://doi.org/10.1007/978-3-642-40585-3_30
Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. Eidsheim, & K. Meizel, K. (Eds.), *The Oxford handbook of voice studies* (pp. 269–300). Oxford Handbooks.
Rothschild, B. (2017). *The body remembers volume 2: Revolutionizing trauma treatment*. WW Norton & Company.
Slatcher, R. B., Vazire, S., & Pennebaker, J. W. (2008). Am "I" more important than "we"? Couples' word use in instant messages. *Personal Relationships, 15*(4), 407–424. https://doi.org/10.1111/j.1475-6811.2008.00207.x
Smith, E. E., Guzick, A. G., Draper, I. A., Clinger, J., Schneider, S. C., Goodman, W. K., Brout, J. J., Lijffijt, M., & Storch, E. A. (2022). Perceptions of various treatment approaches for adults and children with misophonia. *Journal of Affective Disorders, 316*, 76–82. https://doi.org/10.1016/j.jad.2022.08.020
Stalias-Mantzikos, M. E., Barahmand, U., Aguero, K., Peters, L., & Shamsina, N. (2023). Early maladaptive schemas in misophonia.

Clinical Psychologist, 196–209. https://doi.org/10.1080/13284207.2023.2171786

Tannen, D. (1986). Introducing constructed dialogue in Greek and American conversational and literary narrative. *Direct and Indirect Speech, 31*, 311–332. https://doi.org/10.1515/9783110871968

Tomlinson, E. R., Yousaf, O., Vittersø, A. D., & Jones, L. (2018). Dispositional mindfulness and psychological health: A systematic review. *Mindfulness, 9*(1), 23–43. https://doi.org/10.1007/s12671-017-0762-6

Treleaven, D. A. (2018). *Trauma-sensitive mindfulness: Practices for safe and transformative healing*. WW Norton & Company.

Van Dam, N. T., Van Vugt, M. K., Vago, D. R., Schmalzl, L., Saron, C. D., Olendzki, A., Meissner, T., Lazar, S. W., Kerr, C. E., Gorchov, J., Fox, K. C. R., Field, B. A., Britton, W. B., Brefczynski-Lewis, J. A., & Meyer, D. E. (2018). Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation. *Perspectives on Psychological Science, 13*(1), 36–61. https://doi.org/10.1177/1745691617709589

Waldron, E. L. (2017). 'This FEELS SO REAL!' Sense and sexuality in ASMR videos. *First Monday*, *22*(1). https://doi.org/10.5210/fm.v22i1.7282

Walsh, K. M., Saab, B. J., & Farb, N. A. (2019). Effects of a mindfulness meditation app on subjective well-being: Active randomized controlled trial and experience sampling study. *JMIR Mental Health, 6*(1), e10844. https://doi.org/10.2196/10844

Westland, G. (2015). *Verbal and non-verbal communication in psychotherapy*. W.W. Norton & Company.

Wicklund, R. A. (1974). *Freedom and reactance*. Lawrence Erlbaum Associates.

Williams, M., Teasdale, J., Segal, Z., & Kabat-Zinn, J. (2007). *The mindful way through depression: Freeing yourself from chronic unhappiness*. Guilford Publications.

Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLoS ONE, 8*(4), e62397. https://doi.org/10.1371/journal.pone.0062397

Zhu, J., Wekerle, C., Lanius, R., & Frewen, P. (2019). Trauma-and stressor-related history and symptoms predict distress experienced during a brief mindfulness meditation sitting: Moving toward trauma-informed care in mindfulness-based therapy. *Mindfulness, 10*(10), 1985–1996. https://doi.org/10.1007/s12671-019-01173-z