Feasibility of the Fake Phone Call: An iOS App for Covert, Public Practice of Voice Technique for Generalization Training

*Eva van Leer, and †Nick Porcaro, *Atlanta, Georgia, and †Mountain View, California

Summary: Objectives. Generalization is a challenging phase in voice therapy, involving the implementation of a targeted voice technique in all spoken communication. Among other barriers to generalization, self-consciousness keeps patients from practicing and recalibrating their voice technique when they can be overheard (eg, at work). We developed an iOS application that covertly assists users in producing their target voice while they appear to be engaged in a cellular phone call.

Methods. To examine the feasibility of this Fake Phone Call, 11 adults in the generalization stage of voice therapy received a simulated call four times daily for 1 week. Usability, utility, and preliminary efficacy of the application were assessed via triangulated measures including self-report scales, a semistructured interview, and perceptual voice quality assessment of each completed phone call.

Results. Results indicated good feasibility, usability, and utility of the Fake Phone Call in eliciting target voice practice in public without bystander detection of the call's simulated nature. Preliminary efficacy data suggested a positive effect on vocal self-evaluation skill.

Conclusion. The Fake Phone Call and comparable mobile strategies hold potential to facilitate covert public practice and generalization of a speaking-voice technique.

Key Words: Voice therapy−Voice disorders−Treatment adherence−Mobile health−iOS application.

INTRODUCTION

Generalization is a challenging phase of voice therapy. Following the mastery of a prescribed voice technique (eg, confidential, resonant, or loud voice[1,2]) in practice settings such as the clinic, home, or car, the vocal target must be implemented and maintained across communicative settings in the patient's daily life. Patients report a variety of barriers to generalization,[3] including difficulty (1) remembering to use the target voice, (2) establishing the target voice outside of the typical practice environments, (3) attending to both voice production and communication simultaneously, and (4) self-correcting their technique in the midst of a conversation or presentation. In addition, patients avoid practicing or recalibrating their voice technique when others can overhear them,[3] as is typically the case at work, school, or other social settings. As such, patients in the generalization phase may make few generalization attempts during a typical work or school day, fail in the attempts they do make, and persist in a problematic voice mode for the remainder of the day. It is not surprising that a significant number of patients either do not reach treatment goals or fail to maintain these after discharge.[4] Therefore, tools are needed to support patients in the generalization phase of therapy.

Given the present ubiquity of cellular phones,[5] numerous software applications or "apps" have been developed to encourage and support volitional behavior change.[6] These apps range in functionality from basic reminders (eg, calendars, timers) to complex interactive health-monitoring and feedback systems such as the Nike-iPod pedometer.[7] The theory of captology[8]—the study of computers as persuasive technologies—posits that the success of such apps is largely determined by the social acceptability of both the app itself and the behavior it seeks to elicit. For a persuasive technology to trigger a target behavior, both the triggering technology and the targeted behavior must be free from social deviance.[8] For example, individuals are likely to use an unobtrusive app-based pedometer because the behavior of monitoring one's mobile device and the behavior of running are both socially accepted. Conversely, socially deviant behavior and socially awkward technology are unlikely to be used as these draw unwelcome attention. Relevant to voice technique, voice exercises (eg, lip trills, humming, chanting) are perceived by most patients as socially deviant enough to restrict their practice to private settings.[3] To encourage voice recalibration and generalization in social settings, both the elicited behavior—target voice technique practice—and the facilitating technology must be socially acceptable. Today, engaging in cellular phone calls is a common public phenomenon.[9] Therefore, the cellular phone call can potentially serve as vehicle for public voice technique practice without the traditional social deviance and embarrassment.

An iOS app was developed that gives the appearance of a cellular phone call while in actuality engaging the user in target voice practice. This Fake Phone Call app has three aims: (1) to appear as a phone call although it is a voice practice tool, (2) to facilitate target voice production, in particular in public, and (3) to increase target voice use (ie, generalization) after call completion. When engaging with the app, users imitate or elaborate upon utterances presented in the modeled target voice quality, similar to activities completed in a typical voice therapy session. Modeled utterances are scripted to elicit

Accepted for publication February 20, 2018.

From the *Department of Education Psychology, Special Education and Communication Disorders, Georgia State University, Atlanta, Georgia; and the †moForte, Inc. Mountain View, California.

Address correspondence and reprint requests to Eva van Leer, Department of Education Psychology, Special Education and Communication Disorders, College of Education, Georgia State University, 30 Pryor Street Room 808, Atlanta 30303, GA. E-mail: eva.van.leer@gmail.com

Journal of Voice, Vol. 33, No. 5, pp. 659−668
0892-1997

https://doi.org/10.1016/j.jvoice.2018.02.014

the target voice (eg, resonant "wow!" or "um-hum") but restricted to those that are appropriate within the context of a phone call. The task potentially serves as a socially acceptable form of voice technique recalibration.

Although the Fake Phone Call app allows recording of any model (eg, the clinician or the patient), participants served as their own model (ie, the self-as-model) in this feasibility study so that vocal target was unequivocally exemplified in each call. To examine the feasibility of completing both brief and extended Fake Phone Calls, two types of tasks were examined: single word repetition in the "short call" and extemporaneous speech in response to open-ended questions in the "long call."

PURPOSE

The overarching aim of this study was to evaluate the feasibility of using a simulated phone call app—the Fake Phone Call—as a covert practice tool for target voice production. The purpose of this study was threefold:

(1) To assess app *usability*—the ease of use for learning and handling the app[10]—including the ease of feigning a phone call with its support.
(2) To assess app *utility*—the usefulness or appropriateness of the app in achieving its goal[10] of eliciting the user's target voice mode during the phone call.
(3) To obtain preliminary efficacy data regarding the app's effect on generalization and self-efficacy for generalization.

To accomplish these goals, participants in the generalization stage of resonant voice therapy were provided with two types of Fake Phone Calls per day for 1 week's duration. Usability, utility, and preliminary self-efficacy were assessed via a mixed-method approach involving (1) self-report measures at the start and end of the study, as well as following each call, (2) perceptual analysis of voice quality during calls, and (3) a semistructured interview.

METHODS

Data collection took place at the Georgia State University voice laboratory after permission was granted by the Georgia State University Institutional Review Board. All patient instruction, data collection, and data analysis were performed by the principal investigator and two student research assistants.

Research design

Participant perceptions of app usability and utility were obtained after 1 week of daily app use. A quasi-experimental design was applied to compare usability and utility of the two call types: the short and long call. Each call type was presented twice daily. Preliminary efficacy of the Fake Phone Call was assessed with a pretest-posttest design.

Participants

Eleven adults agreed to participate in this study, including nine women ages 22–56 years (mean = 30.13 years, standard deviation [SD] = 11.05 years) and three men ages 24–28 (mean = 26.3, SD = 2.52 years). Three participants identified themselves as African American and eight as Caucasian. As part of a larger study, all participants had received four sessions of voice therapy to address vocal hyperfunction[11] with or without laryngeal consequences noted in Table 1. Having demonstrated in-session achievement of resonant voice at the sentence level in this preceding study, participants were in the generalization stage of therapy upon entering the present study.

TABLE 1.
Participant Vocal Fold Status

Etiology	n
Normal	2
Normal status post thyroidectomy	1
Bilateral vocal fold edema and erythema	5
Mid-membranous polyp, unilateral	1
Ectasia, mid-membranous edema	1
Vocal fold bowing	1

iOS application

An iOS software app was created in Objective C (a programming language). In this system, a functionality was developed that works as follows. First, a sound file is recorded at 44K sampling rate in wav format. This sound file contains model utterances and cues that will prompt the user at the time of responding to the call. The sound file is tagged as "fake phone call" so that future playback will occur via the earpiece instead of the speaker when the app is installed on an iPhone. Reminders are set for the times of day each recording will be played. At the time of the reminder, the user receives a notification which plays a telephone ring. The notification appears to the user whether the app is running or not and gives the illusion that the user is receiving a phone call. When the user "picks up the phone" by swiping the notification on the screen, the prerecorded sound file plays via the earpiece. The user's responses are recorded, yielding a new sound file in which responses are intermingled with the prerecorded utterances. The phone call is followed by a brief survey asking the user to rate their voice quality before and during the call, as well as the awkwardness of completing the call on a 0–10 scale. The sound file is saved and an entry is logged in the database regarding the duration and time of call, and survey answers. These are exported for analysis.

Patient instruction and data collection protocol

Participants visited the laboratory on two occasions, 1 week apart. On the first visit, self-report measures were completed, the app was installed on participants' personal iPhone, and participants were instructed regarding its use, including swiping to answer the call and completing the brief on-screen survey after call completion. One participant loaned an iPhone from the laboratory. Next, participants were coached to achieve their optimal target voice

production mode. Short and long phone call scripts were practiced and recorded into the app in the participants' target voice mode.

For the recording of the self-as-model short call, an individualized script was made in collaboration with each participant. The script consisted of approximately six brief one to two-word utterances that best elicited resonant voice for that particular participant and were also plausible in a phone call context, such as "Hey you!" "Wow, wow!" "Mhm, mhm." For example, one participant's short call script was "Hey, what's up? Yeah. Yeah, yeah, yeah! Hmm. Hm-m? OK. Alright. Alright, bye." Following each utterance, an approximately equal amount of silent time was recorded to repeat the utterance upon receiving the call. This yielded a recording of short words and phrases alternated with equal periods of silence.

For the long call, the participant was recorded asking three open-ended questions: (1) "How was your morning?", (2) "Any plans for this afternoon?", and (3) "What are you doing for dinner?" Participants were allowed minor variations. For example, one participant's long call script was "How are you?", "How is school going?", and "What are you doing for lunch and dinner?" After each question, a 10-second silence was recorded for production of an extemporaneous answer upon receiving the call. After the final (third) silence, the call was completed with the utterance "Bye!" to signal the end of the call.

In the app's reminder menu, one short call alert was scheduled for the midmorning and another for the midafternoon. Similarly, long call reminders were scheduled for the late morning and afternoon. Notification times were individualized to optimize the likelihood that the participant would be available at the time of each call. Participants were instructed to respond to as many scheduled calls as possible and to complete the subsequent on-screen survey. There was no option to return a call if a call was missed.

On the second visit at completion of the study, the resulting intermingled sound files of answered calls and the post-call survey data files were exported from the participants' phone. These were stored on the laboratory computer for future analysis. Prestudy questionnaires were completed again, and a semistructured interview was conducted regarding the experience and subsequently transcribed.

MEASURES

System Usability Scale

Participants completed the System Usability Scale[12] (SUS) at the end of the 1-week study duration to assess usability. The SUS is a global, widely employed, subjective assessment of usability that is validated and highly reliable (alpha = 0.91).[12] The SUS comprises 10 items rated with a Likert-style 5-point response format ranging from 1 (strongly disagree) to 5 (strongly agree). Scoring is completed by subtracting 1 from odd-numbered (positively worded) responses and subtracting the even-numbered negatively worded responses from 5. SUS scores range from 0 to 100 but do not reflect percentile ranks. Sixty-eight points represent the average usability score across systems.[13] Total scores also were found to correspond with seven qualitative adjective ratings,[14] ranging from "worst imaginable" at the low end to "best imaginable."

Semistructured interview

A semistructured interview was conducted on the second visit (ie, at the end of a week of app use) to capture participants' perceptions of both usability and utility. The following open-ended questions were asked: (1) "What was it like to use this app?", (2) "What was it like to receive the short call?", (3) "What was it like to receive the long call?", (4) "How awkward was it to complete these fake calls?", (5) "What was it like to do so with people around you?", (6) "For whom or when in therapy would this app be useful?", and (8) "How would you improve or change the app?" Questions were adjusted in response to participants' answers to capture perspectives thoroughly.

On-screen survey

To enhance the poststudy semistructured interview with participant perspectives at the time of each call, a brief survey appeared on the iPhone screen following each call. The survey employed a 0−10 response format. To assess usability, participants rated the awkwardness of each phone call, with 0 representing no awkwardness and 10 representing maximum awkwardness. To assess utility, participants were asked to separately rate their "voice quality" preceding the call and during the call, with 0 representing their problematic habitual voice production mode, and 10 representing their optimal "on-target" voice used in the model recordings. Lastly, participants were asked to rate their "confidence" for continuing to use the target voice after call completion, comparable to their self-efficacy ratings in the Readiness Ruler and Self-Efficacy Scale for Voice Therapy.

Perceptual discrimination of voice quality

To determine utility of the app in eliciting the target voice quality in responses to the calls, all completed calls were judged perceptually by two graduate research assistants in the manner of a discrimination task. Participants' voice quality in call responses was judged in reference to their target voice quality demonstrated in their prerecorded scripted utterances. An "on-target" rating was assigned when voice matched or exceeded the target, whereas an "off-target" judgment was assigned when response voice quality was not as good as the target. Recordings were presented at comfortable loudness over Sennheiser HD205 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany).

Before analysis of the entire corpus of calls, three research assistants were screened for their ability to discriminate on-target from off-target productions. Ten calls exemplifying on- and off-target productions in both short and long calls were chosen by the first author, including three off-target and two on-target short calls, and the same distribution of off- and on-target long calls. In making perceptual discrimination judgments independently, two research assistants agreed with the author for 9 out of 10 calls and were therefore assigned the

task of rating all calls after a 1-week washout period. A third rater agreed on only 4 of 10 occasions, and was therefore excluded from perceptual data analysis.

Generalization: consistency

On both visits, participants were asked to estimate the percent of time they used their target voice over the course of the preceding week, ranging from 0% (not at all) to 100% (all the time). This approach was previously used in our research.[15]

Generalization: accuracy

As the target voice mode was associated with the absence of laryngeal effort, on both visits, participants were asked to rate their habitual laryngeal effort in the past week on the Borg CR10 scale adapted for vocal effort.[16] The scale employs a numeric response format ranging from a value of 0−10, as well as half points (eg, 0.5). The numeric points are anchored by categorical verbal expressions including, for example, "very, very slight effort" commensurate with a numeric score of 0.5, to "severe vocal effort" commensurate to a score of 5. The scale end points were experientially anchored with the descriptions "how your larynx feels when you're completely in resonant voice, like in humming" for the 0.5 point and "how it feels when you're trying to talk through bad laryngitis" for the value of 10 (ie, maximal vocal effort).

Self-efficacy for generalization

On both visits, participants rated their confidence in their own ability to use the target voice throughout the day (ie, self-efficacy for generalization) with item 1 of the Readiness Ruler,[17] as well as with the generalization subscale of the Self-Efficacy Scale for Voice Therapy.[15] Readiness Ruler item 1 solely asks participants to rate their self-efficacy for using the target voice mode consistently, thus representing an overall estimate of self-efficacy for generalization. Alternately, the generalization subscale of the Self-Efficacy Scale for Voice Therapy asks participants to rate this same construct in the context of 17 challenging situations (eg, when talking over noise, when you're busy, when talking to people who are unsupportive of your voice problem), each represented by a scale item. An average score is calculated. Both scales employ a 0- to 10-point response format ranging from "not at all confident" to "extremely confident" to use the target voice.

Statistical analysis

All analysis was completed using IBM SPSS Statistics 20 (SPSS Inc., IBM Corp., Armonk, NY). For purpose 1, only descriptive analysis was needed. Also, as an exploratory measure, a Pearson correlation was calculated to quantify the relationship between the number of calls completed and the SUS score. For purpose 2, the proportion of calls rated "on target" was calculated for each call type and rater. Subsequently, these proportions were averaged across the two raters for each call type, yielding an averaged percent of accurate short calls and an averaged percent of on-target long calls. A binomial test was completed to assess whether recordings were on target at a probability greater than chance. A chi-square and Fisher exact test procedure was conducted to assess whether these proportions differed significantly by call type as this would indicate participants performed better on one call task than another. Furthermore, paired *t* tests were completed to examine whether differences in the mean were significant for patient self-ratings of voice technique during the call in the on-screen survey and for all measures completed before and after study for purpose 3. In addition, for exploratory purposes, the association between generalization and each measure of self-efficacy was calculated using a Pearson correlation.

Qualitative analysis

Transcripts of semistructured interviews were analyzed for content categories, sub-themes, and themes, as consistent with standard qualitative methods.[18] Two raters—the principal investigator and a research assistant—independently read and coded the transcripts, and subsequently organized these into responses regarding the overall utility and usability of the call, and perceptions of each call type.

RESULTS

Usability

Quantitative results

Participants completed a mean of 7.18 short calls (SD = 2.44) and 6.36 long calls (SD = 3.38) over the course of 1 week, corresponding to approximately one short and one long call per day. The mean SUS score was 74.5 (SD = 7.57). This exceeds the average published SUS score of 68 for mobile systems[13] and corresponds to a patient-perceived adjective rating of "good."[14] Awkwardness ratings on the postcall survey were low, with means of 2.01 (SD = 2.4) out of 10 for the short call and 1.84 (SD = 1.9) for the long call. Differences between these means were not significant. There was a significant and moderately strong association between SUS scores and number of completed calls as indicated by $r = 0.651$, $P < 0.029$. Thus, individuals who perceived the app as more usable also completed more calls.

Qualitative results

Overall, usability was perceived as good. Three themes emerged from interviews regarding the underlying mechanism of good usability: easy to use, easy to feign, and barely awkward. First, the system was perceived as technically easy to use. Only one participant—an android phone user who borrowed a laboratory iPhone for the study—experienced an initial technical difficulty related to swiping the notification banner. Second, the simulated nature of the call was easily concealed. Participants denied bystander detection of the calls' simulated nature. Two female participants were asked to reveal the caller's identity (eg, "Who were you talking to?") by their boyfriend and father, respectively. Third, the awkwardness associated with completing the call was perceived as low by all participants, as also noted quantitatively in the low on-screen awkwardness ratings. Several participants reported mild awkwardness related to (1) listening to one's own voice (n = 2) ("I hate my voice"), (2) the

TABLE 2.
Usability Themes, Subthemes, and Example Statements

Theme/Category	Subtheme, Content Codes	Example Statements
Easy to use	Technically easy to use	"It worked extraordinarily well."
Easy to feign	Easy to feign simulated nature	"No one noticed."
Barely awkward	Low awkwardness	"Other people aren't going to know, so if it's awkward, you're the one making it awkward. The system is not awkward."
		"Getting a call from yourself is weird."
		"When I was totally alone it was most awkward talking to myself by myself."
Innovations: flexibility	Callback function	"When I missed a call, I couldn't do my fake phone call. There was no way to "call back."
	Interactive timing for long call	"I'd get "cut off" by the next question."
	Call script database	"...awkward silences fell"
	Set number of calls yourself	"After I did it a few times I got bored so I would just respond in different ways."
		"I felt guilty when I didn't feel like taking a call; the way you feel guilty when you're avoiding a friend who calls too much."
Innovation: feedback	Review/playback function	"It was hard to judge if my voice improved without being able to listen to it."
	Objective feedback	"I wanted to know if it was, like, spectrally correct."

unnatural experience of getting a call from oneself (n = 1), and (3) feigning a call when there were no observers for whom to feign the simulated nature (n = 1). Participants reported these as minor issues they had to "get over." Example statements are found in Table 2.

When asked how the app may be improved, two themes emerged: flexibility and feedback. To increase the flexibility of the simulation, making it more like a real phone call, the following development suggestions were made: (1) a "callback" option to answer missed phone calls (n = 3), (2) a larger number of call types or scripts to avoid boredom with the task (n = 2), (3) ability to adjust the number of calls per day: fewer calls were requested by one participant: "I felt guilty ignoring calls, the way you feel guilty when you ignore a friend who calls too often," and (4) voice detection to improve timing of questions in the long call to avoid both "awkward silences" and being "cut off" (n = 2). With regard to feedback, one participant requested a playback function to review one's recording before completing the on-screen survey ("I wanted to hear how I did") and another suggested objective acoustic analysis of performance.

Utility results

Perceptual voice quality

A total of 148 calls were completed: 78 short calls and 70 long calls. In perceptual evaluation, inter-rater agreement was 0.94 for reliability in judging voice quality dichotomously as "on target" or not compared to the self-as-model anchor. Averaged across the two raters, 77% of short call responses and 66% of long call responses were perceptually rated as on-target. Although a greater percentage of short calls than long calls were on target, the difference in proportions was not significant as indicated by Fisher exact test ($P = 0.14$). The binomial probability that on-target voice was achieved was significantly greater than chance for both the short call ($P = 0.00000001$) and the long call ($P = 0.0029$).

Self-rated technique in the on-screen survey

On average, participants retrospectively rated their voice quality as significantly better during the call than immediately before answering the call, for both the short and long call (Table 3). For the short call, 72 postcall surveys were submitted by the group in total. There was a significant improvement in self-perceived voice technique ratings as calculated with a paired t test, where $t(71) = 6.82$, $P = 0.00000000124$, and Cohen effect size was large at 0.8. For the long call, 59 ratings were supplied, and mean increases in self-perceived vocal improvement were again significant, with $t = 5.71$, df = 58, $P = 0.000000204$; again Cohen effect size was moderately large at 0.74. Although participants rated their voice technique better in the short than in the long call, the difference between means was not significant.

Semistructured interviews: utility

Participants spoke positively about the utility of the call in eliciting target voice production. Four themes emerged regarding the mechanism of this utility. Details are provided in Table 4.

TABLE 3.
Self-Rated Voice Quality Before and During Calls

Call Type	Before Call Mean (SD)	During Call Mean (SD)	Mean Difference
Short call	6.38 (2.29)	8.07 (1.31)	1.74 (2.12)
Long call	6.83 (1.68)	7.93 (1.36)	1.12 (1.49)

TABLE 4.
Utility Themes, Content Codes, and Example Utterances

Theme/Category	Subtheme, Content Codes	Example Quotation
Helps me remember	Reminder reduces cognitive demand	"It took some of the mental work away out of having to remind myself to stay in my good voice."
	Reminder might help you return to target voice	"I think with voice it's so much more subtle (than remembering physical exercises) so it takes a lot more mental energy because you're so used to your habits."
		"Morning and midday were important. . .to start my day with where my voice needs to be, and (midday) to get back to where I need to be."
		"Getting a text could be enough for people with good generalization already but not others."
Physical practice is vital	Engage in voice production	"You have to go through the process of answering the call and listening to it and it has more of an impact."
	Accountability/maintenance	"Because you have to practice. It's like hands-on experience versus reading about something. You can't just read about it, because then it's just like making empty conjecture."
		"Having to do the calls made me accountable for maintaining my therapy voice."
Vocal awareness	Calls raised awareness about generalization accuracy and consistency	"It was humbling. I thought I was always in my good voice, but the calls helped me realize I'm not as consistent as I thought."
		"No one is perfect all the time. So it was good to have."
		"There's still stuff that could be more bright and clear; I should (be back in therapy) so it becomes more natural. I've actually kind of forgotten what I need to be doing: I can remember it kind of on a whim but I need an actual routine, I need to practice."
Matching yourself (is useful)	Self-as-model is a good target	"It reminded me of what I could be aiming for."
	Self-as-model increases self-efficacy	"I was definitely trying to match myself. . .I felt I was pretty successful."
		"It reminded me that I could have a strong voice, it was reassuring that I could do it right."
Matching yourself (is weird)	Self-as-model is a difficult target because it is unfamiliar	"It's more difficult to model my voice than it is to model someone else's. I never had an experience where it clicked for me to (listen to) my voice."
		"Hearing myself was not unpleasant; however, being able to mimic myself was harder than mimicking someone else."

Helps me remember. The banner notification of the call accompanied by the classic ring tone served as reminder to implement the target voice technique even when a call could not be answered at that time. Participants reported that this memory aid alleviated some of the mental effort of remembering to implement the target voice. One participant felt that, toward the end of the study, she could return to target voice mode simply by viewing the reminder, but before this point she and all other participants required participation in call completion to recalibrate to the target voice mode.

Physical practice is vital. All participants emphasized the importance of physically engaging in the call to achieve target voice or progress toward this goal. All elements of physical practice were listed as important: listening to the model, producing voice in reference to the model, evaluating one's performance for each utterance, and repeating the cycle with the next utterance. Participants noted that they often failed to reach the target voice (ie, a 10 in the on-screen survey) and had to continue focusing on their technique after call completion. Speaking out loud was perceived as vital to finding the target.

Vocal awareness. Participants reported that completing the calls increased their ability to self-evaluate their voice technique, and thus, their ability to judge the accuracy and consistency of generalization. Hence, several participants felt "humbled" when they realized they were not as consistent or as accurate as they had estimated before the study. Relatedly, all participants thought the calls could be useful for patients in the generalization and maintenance phase, but two proposed incorporating them early in therapy to raise vocal awareness of poor habitual voice use. One "humbled" participant abandoned the goal of vocal improvement at study completion, stating that "I'm not willing to do what it takes to get there" (ie, to consistent target voice use).

TABLE 5.
Short and Long Call Themes, Content Codes, and Example Statements

Theme/Category	Subtheme, Content Codes	Example Statement
Shorter call is easy	Motor skill is easy	"It was helpful especially on the "yeah yeah yeah's" because that was what I was supposed to do to get better and I did improve and get better."
	Feigning easy for short time	"No one noticed. That's why I liked the short call; it was very covert, very stealth."
	Short duration is convenient	"If I was rushed, I could do it really quickly."
		"It was good in that there was less phrases so it was easier to get done there."
Duration matters	Too little practice in short call	"But I found that it needed to be longer and that's where the long call would come in."
	More practice in long call	"I like the longer better because you can really monitor what you're doing, instead of the short call where the short phrases, you can easily nail that, but does it really help you?"
		"In terms of talking in conversation the short one was not enough."
Long calls are hard	Generating sentences is awkward in fake call context	"It was more awkward because I had to come up with something to say, like 'I'm going to the store' and it almost made it weird to whoever I was with."
	Motor task and dual task more difficult	"You had to stay in your good voice."
		"You had to think of something to say."

Matching yourself is useful. Participants reported utility of the self-as-model, listing that this example (1) presented a goal "to aim for," (2) reminded them that target voice use was possible, and (3) increased awareness of generalization by allowing comparison to the participant's voice at the time of the call.

Matching yourself is weird. Two participants experienced matching yourself as problematic because this task was novel (the clinician model had primarily been used in therapy) and unusual in call context.

Comparison of short and long call

Short call is easy. The short call was perceived as "easy' because it was brief, requiring a short period of feigning, and because difficulty level was low (ie, short resonant utterance repetition). However, as it was easy to "nail it" from a motor learning perspective, one participant doubted the efficacy of the short call in attaining resilient resonant voice after call completion.

Long call is hard. The long call was perceived as more difficult for several reasons: (1) a longer duration of feigning was required, (2) the vocal task—extemporaneous speaking in resonant voice—was perceived as much harder than the word-level repetition of the short call, and (3) the task required cognitive effort to "think of something to say."

Duration matters. The long call was considered useful precisely because it allowed more time to "find the (resonant) voice" and better resembled actual conversational voice demands.

Example statements related to all themes are noted in Table 5.

Preliminary efficacy data

Preliminary efficacy data means are shown in Table 6. Borg CR10 effort ratings reduced significantly from before to after the study, as indicated by paired t test results of $t(10) = 1.84$, $P = 0.048$. As such, there was a significant and moderate correlation between the number of calls completed and the reduction in Borg scores, with $r = -0.6274$, R^2 at 0.39, $P = 0.0387$. However, there were no significant

TABLE 6.
Pre- to Poststudy Outcome Measures

Measure	Study Onset M (SD)	Study Completion M (SD)
Borg CR10 Vocal Effort Score	2.64 (1.39)	1.68 (.98)
Overall Self-Efficacy for Generalization (Readiness Ruler item 1)	7.81 (2.4)	7.72 (1.27)
Self-Efficacy Scale for Voice Therapy Mean (generalization subscale)	6.97 (2.03)	7.21 (1.38)
% of Time generalization (ie, using target voice in past week)	59.09 (28.71)	69.09 (20.71)

differences between the remaining pre- and poststudy outcome measures. Descriptive data are noted in the table below. Interestingly, all five participants with low initial generalization levels (ie, below 70%) reported higher levels of generalization at the end of the week, whereas those with high initial levels of generalization had variable self-ratings at study completion, indicating no change, improvement, or reduction of generalization. Mean self-efficacy ratings obtained in the on-screen for the short call (8.34, SD = 0.68) was significantly higher than poststudy self-efficacy ratings obtained with the Self-Efficacy Scale for Voice Therapy ($t(10) = 4.02$, $P = 0.001$) and Readiness Ruler ($t(10) = 1.86$, $P = 0.046$). Likewise, mean self-efficacy ratings after the long call (7.97, SD = 1.04) were significantly higher than Self-Efficacy Scale for Voice Therapy ratings ($t(10) = 1.88$, $P = 0.04$) but not significantly higher than Readiness Ruler ratings. Thus, participants largely felt more confident immediately after receiving Fake Phone Call support than when rating self-efficacy outside of the call context.

Exploratory analysis

Because participants reported that increased vocal awareness improved their ability to assess their own generalization, their self-efficacy ratings for generalization may have improved in accuracy as well. A "humbling effect" would lower the traditionally higher (and possibly overestimated) self-efficacy judgments of the Readiness Ruler, thus improving predictive validity of the Readiness Ruler. An exploratory analysis was therefore conducted to examine the effect of study participation on the predictive validity of the self-efficacy scales. For the Self-Efficacy Scale for Voice Therapy, scores were strongly and significantly associated with generalization both at study onset (Spearman R = 0.851, $P = 0.0009$) and at study completion (R = 0.889, $P = 0.0024$) as consistent with our previous research demonstrating good predictive validity of this scale. However, for the Readiness Ruler, predictive validity increased from a moderate association with generalization at study onset (Spearman R = 0.69, $P = 0.0187$) to a strong association at study completion (R = 0.898 and $P = 0.00017$).

DISCUSSION

The overarching aim of this study was to assess the feasibility of covert public voice technique practice via the Fake Phone Call, an iOS app that engages the user in target voice practice via a simulated cellular phone call. Triangulated interview, self-report, and behavioral data yielded three key findings: (1) usability of the app was good with regard to both technical and social-behavioral factors, (2) both the short and long call tasks held utility for eliciting the target voice mode, but differed somewhat in their mechanism of usefulness, and (3) accuracy in self-assessment of target voice production appeared to improve as a function of Fake Phone Call completion.

Feasibility of Fake Phone Call completion was intricately related to usability. With the traditional telephone ringtone and playback via the earpiece as well as the conversational target voice practice task, the illusion of a phone call was successfully upheld in public for an average of two calls a day. Qualitative themes "easy to use," "easy to fake," and "barely awkward" were supported by quantitative data of good SUS scores and low awkwardness ratings in the on-screen survey. The intricate relationship between technical and social-behavioral usability factors was further illustrated by the significant association between SUS scores and the number of calls completed, as well as the usability preference for the short call. Taken together, findings exemplify the primacy of social acceptability in captology.[8,19]

App utility in eliciting the target voice quality was demonstrated in perceptual analysis results, patient self-ratings of vocal improvement in the on-screen survey, and patient perspectives reported via qualitative interview data. The mechanism of this utility relates to several known processes involved in meta-cognition,[20,21] principles of motor learning,[22] and goal-directed behavior change.[23]

The utility of the notifications captured in the theme "helps me remember" addresses the known barrier of "forgetting" to use the target voice, documented in our previous research.[3] The usefulness of call notification is not surprising, given that electronic memory aids have demonstrated efficacy to support memory-dependent activities in everyday life and counter self-regulatory fatigue.[20,21] However, although reminders were useful, the primary utility of the app lay in engaging in calls, as indicated by a triad of qualitative themes: physical practice is vital, matching yourself, and vocal awareness.

The theme "physical practice is vital" speaks to the challenge of generalizing a target voice mode to different environments and novel utterances in the process of motor learning. As stated by Verdolini, one cannot simply "summon" a desired voice production mode across environments[22] when that mode is not habitual. Planned target voice use is needed across communication environments to generalize and habituate the desired vocal mode. Clinical research regarding, for example, Lee Silverman Voice Therapy training,[2] has shown benefit of structuring such generalization activities as part of the treatment protocol. The Fake Phone Call provides scaffolded support for this process in engaging the user in the target mode across environments in a communication-like activity.

The reported utility of "matching yourself" lay in the concreteness of the vocal target to "aim" for. The theme relates the known motivational mechanism of goal characteristics laid out in goal setting theory.[23] Concrete goals have shown to be to be inherently more motivating than abstract, vague goals, precisely because concrete goals allow one to judge his or her progress in relation to the goal.[23] In our previous research, patients described resonant voice targets as "abstract" and "vague" due to their kinesthetic and nonvisual nature.[3] The documented barrier of "not knowing if you're doing it right" was addressed in the present study by incorporating the self-as-model as vocal target. Positive responses to the self-as-model were also found in our prior work involving the video self-as-model.[15]

Unexpectedly, the task of matching yourself also appeared to improve vocal self-assessment skill, in particular when paired with self-assessment in the on-screen survey. This resulted in an entirely separate theme: vocal awareness. Therefore, the Fake Phone Call holds potential as a vocal self-evaluation or awareness tool. In relation to research outside of voice science, self-assessment tasks have shown to be useful for a variety of speech therapy tasks, including fluency training.[24] The challenge in voice therapy may lie in increasing awareness without demoralizing the patient and causing dropout secondary to the "humbling effect."

The Fake Phone Call was designed not only to trigger the target voice mode, but to encourage its generalization beyond the call context. Preliminary efficacy data yielded mixed results toward this aim. On average, pre- and post-study generalization scores did not differ significantly, nor did associated self-efficacy for generalization, although self-efficacy scores were elevated following each call, as indicated by on-screen survey values. Borg CR10 scores, however, were significantly reduced by study completion, indicating increased accuracy of target voice use (ie, reduced hyperfunction). Taken together, the two findings offer preliminary evidence to suggest that Fake Phone Calls improve the accuracy but not the frequency of target voice use. The research design does not allow us to conclude that Fake Phone Call practice is more effective than unassisted practice and generalization, because control conditions were not incorporated in this feasibility trial.

Although all poststudy generalization scores may have been restricted by the humbling effect, a closer examination of the generalization data shows that all five participants with low initial generalization scores (ie, below 70%) rated their generalization as higher at study completion, whereas those with high initial scores (ie, 70% or above) either varied or plateaued in their generalization at study completion. These data may indicate that efficacy of the Fake Phone Call is limited to patients who have not yet achieved high generalization or are more modest in their generalization estimates. This finding resonates with our previous studies showing benefit of mobile support for individuals with low self-efficacy for practice, but not for those with initial high self-efficacy.[15,25]

The humbling effect of vocal self-awareness was also evidenced in self-efficacy data. Self-efficacy measures are known to be more predictive of actual behavior (eg, practice and generalization) when the construct of self-efficacy is assessed in relation to specific barrier situations, each represented by a different scale item.[26] The more complete and detailed the list of barrier situations, the greater the predictive validity of the scale.[26] In the absence of barrier items, individuals overestimate their self-efficacy for a task, resulting in high scores that are less predictive of behavioral measures. This was confirmed in our previous studies: the multi-item Self-Efficacy Scale for Voice Therapy had greater predictive validity than the single self-efficacy item on the Readiness Ruler.[15] However, at completion of the present study, Readiness Ruler means were lower and more predictive of generalization than at study onset, behaving comparably to multi-item Self-Efficacy Scale for Voice Therapy. The finding suggests that a single-item measure of self-efficacy has the potential to hold predictive validity, but only if a patient's vocal self-awareness is good.

LIMITATIONS

Our study was limited by several factors. First, all participants were familiar with cellular phones and apps. Inclusion of a greater number of participants with limited exposure to iOS devices and apps could have yielded different results, in particular for usability. Second, findings are limited to individuals with mild to moderate hyperfunctional dysphonia, utilizing the app in the generalization phase of a resonant voice therapy approach. User experiences might vary depending on dysphonia severity, stage of therapy, and type of therapy. Third, only self-as-model recordings were examined within the Fake Phone Call context, although the system can allow any type of model uploads, such as clinician model examples. Clinician models would likely result in different utility experiences. Lastly, a full efficacy study was not completed, as this would have required sampling of the target voice throughout the day for comparative baseline data.

Clinical Implications

As an applied treatment study, key findings relate directly to the clinical setting, within the discussed limitations. Because faking a call was found to be surprisingly feasible, doing so has the potential to increase practice frequency by allowing public practice in addition to traditional home or car-based practice. The theme of matching yourself suggests that the self-as-model may hold utility in and of itself (ie, outside of the Fake Phone Call) by increasing goal concreteness. The greater perceived difficulty of the long call speaks to the challenge of generalization, because the long call remains a much easier task than real-world generalization itself. With an average treatment duration of 4.2 sessions,[27] it is unlikely that patients have achieved consistently accurate target voice use at discharge. Improved patient vocal awareness could result in more accurate self-report of adherence, with important consequences for treatment and discharge planning.

The theme of physical practice has implications for the utility of wearable ambulatory phonation monitors and feedback systems.[28] Wearable systems provide covert feedback based on the patient's natural conversational voice use, without requiring a dedicated task: a great benefit compared to the Fake Phone Call, which could not be completed on half of the scheduled occasions. However, covert feedback systems do not provide a modeled practice opportunity perceived as "vital" by our participants. Thus, a higher skill level may be needed to benefit from wearable feedback: one in which the patient can "reset" their voice on the basis of feedback alone, without engaging in modeled physical practice. A combination of ambulatory feedback and Fake Phone Calls may provide maximum support for generalization.

Future research

The present study represents a proof of concept of the Fake Phone Call. Now that feasibility is established with a good understanding of usability and utility, future directions involve both efficacy testing and further app development. As an immediate next step in app development, we will create a "callback" functionality for cases in which the user has missed a phone call. A long-term goal is to develop a performance feedback system, because knowledge of results is necessary for motor skill acquisition and learner motivation.[23,29]

CONCLUSION

Feasibility, usability, and utility of the Fake Phone Call app were demonstrated for covert public practice of a vocal technique. Study results shed light on the potential for mobile tools to reduce practice and generalization barriers in voice therapy, and on the utility of the self-as-model in voice therapy.

Acknowledgments

This work was supported by the American Speech and Hearing Association (ASH) Foundation Clinical Research Grant and by the National Institutes on Deafness and Other Communication Disorders (Grant 1R03DC013884-01). The authors are indebted to Mosheh Wolf for creative discourse and Robert Clayton Hendrick for statistical support.

REFERENCES

1. Verdolini-Marston K, Burke MK, Lessac A, et al. Preliminary study of two methods of treatment for laryngeal nodules. *J Voice*. 1995;9:74–85.
2. Ramig LO, Fox C, Sapir S. Parkinson's disease: speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Semin Speech Lang*. 2004;25:169–180 (Copyright\copyright 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA).
3. van Leer E, Connor NP. Patient perceptions of voice therapy adherence. *J Voice*. 2010;24:458–469.
4. Speyer R. Effects of voice therapy: a systematic review. *J Voice*. 2008;22:565–580.
5. Rainie L. Cell phone ownership hits 91% of adults. *Fact Tank News Numbers*. 2013;6.
6. Boulos MNK, Brewer AC, Karimkhani C, et al. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform*. 2014;5:229.
7. Conroy DE, Yang C-H, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *Am J Prev Med*. 2014;46:649–652.
8. Fogg BJ. *A behavior model for persuasive design*. In Proceedings of the 4th International Conference on Persuasive Technology 40 (ACM). 2009.
9. Rainie L, Zickuhr K. *Americans' views on mobile etiquette"*. (Pew Research Center). 2015.
10. Grudin J. Utility and usability: research issues and development contexts. *Interact Comput*. 1992;4:209–217.
11. Hillman RE, Holmberg EB, Perkell JS, et al. Objective Assessment of Vocal Hyperfunction. *J Speech Lang Hear Res*. June 1989;32:373–392. https://doi.org/10.1044/jshr.3203.373.
12. Brooke J. SUS—a quick and dirty usability scale. *Usability Eval Ind*. 1996;189:194.
13. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. *Int J Hum Comput Interact*. 2008;24:574–594.
14. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud*. 2009;4:114–123.
15. van Leer E, Connor NP. Predicting and influencing voice therapy adherence using social-cognitive factors and mobile video. *Am J Speech Lang Pathol*. 2015;24:164–176.
16. van Leer E, Van Mersbergen M. Using the Borg CR10 physical exertion scale to measure patient-perceived vocal effort pre and post treatment. *J Voice*. 2017;31:e19–e25.
17. Miller WR, Rollnick S. *Motivational Interviewing: Helping People Change*. Guilford Press; 2012.
18. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. 2004;24:105–112.
19. Fogg BJ. *Persuasive computers: perspectives and research directions*. In Proceedings of the SIGCHI Conference on Human factors in Computing Systems 225−232 (ACM Press/Addison-Wesley Publishing Co.). 1998.
20. Tobias R. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. *Psychol Rev*. 2009;116:408.
21. Vinney LA, Turkstra LS. The role of self-regulation in voice therapy. *J Voice*. 2013;27:390 e1.
22. Verdolini K. Principles of skill acquisition applied to voice training. *Hampton M Acker B Vocal Vis Views Voice By*. 1997;24:65–80.
23. Locke EA, Latham GP. *A Theory of Goal Setting & Task Performance*. Prentice-Hall, Inc; 1990.
24. Ingham RJ. The effects of self-evaluation training on maintenance and generalization during stuttering treatment. *J Speech Hear Disord*. 1982;47:271–280.
25. van Leer E. *Utility and usability of an iOS voice quality feedback application*. 2013.
26. Bandura A. Guide for constructing self-efficacy scales. *Self-Effic Beliefs Adolesc*. 2006;5.
27. Portone-Maira C, Wise JC, Johns MM, et al. Differences in temporal variables between voice therapy completers and dropouts. *J Voice*. 2011;25:62–66.
28. Mehta DD, Van Stan JH, Zañartu M, et al. Using ambulatory voice monitoring to investigate common voice disorders: research update. *Front Bioeng Biotechnol*. 2015;3.
29. Wulf G, Shea CH, Matschiner S. Frequent feedback enhances complex motor skill learning. *J Mot Behav*. 1998;30:180–192.