Disability and Rehabilitation: Assistive Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iidt20

Differences of people with visual disabilities in the perceived intensity of emotion inferred from speech of sighted people in online communication settings

Hyung Nam Kim & Shaniah Taylor

To cite this article: Hyung Nam Kim & Shaniah Taylor (2022): Differences of people with visual disabilities in the perceived intensity of emotion inferred from speech of sighted people in online communication settings, Disability and Rehabilitation: Assistive Technology, DOI: 10.1080/17483107.2022.2114555

To link to this article: https://doi.org/10.1080/17483107.2022.2114555

Published online: 23 Aug 2022.

DISABILITY AND REHABILITATION: ASSISTIVE TECHNOLOGY
https://doi.org/10.1080/17483107.2022.2114555

ORIGINAL RESEARCH

Differences of people with visual disabilities in the perceived intensity of emotion inferred from speech of sighted people in online communication settings

Hyung Nam Kim and Shaniah Taylor

North Carolina A&T State University, Greensboro, NC, USA

ABSTRACT

Purpose: As humans convey information about emotions by speech signals, emotion recognition *via* auditory information is often employed to assess one's affective states. There are numerous ways of applying the knowledge of emotional vocal expressions to system designs that accommodate users' needs adequately. Yet, little is known about how people with visual disabilities infer emotions from speech stimuli, especially *via* online platforms (e.g., Zoom). This study focussed on examining the degree to which they perceive emotions strongly or weakly, i.e., perceived intensity but also investigating the degree to which their sociodemographic backgrounds affect them perceiving different intensity levels of emotions when exposed to a set of emotional speech stimuli *via* Zoom.

Materials and methods: A convenience sample of 30 individuals with visual disabilities participated in zoom interviews. Participants were given a set of emotional speech stimuli and reported the intensity level of the perceived emotions on a rating scale from 1 (weak) to 8 (strong).

Results: When the participants were exposed to the emotional speech stimuli, *calm, happy, fearful, sad, and neutral*, they reported that neutral was the dominant emotion they perceived with the greatest intensity. Individual differences were also observed in the perceived intensity of emotions, associated with sociodemographic backgrounds, such as health, vision, job, and age.

Conclusions: The results of this study are anticipated to contribute to the fundamental knowledge that will be helpful for many stakeholders such as voice technology engineers, user experience designers, health professionals, and social workers providing support to people with visual disabilities.

ARTICLE HISTORY
Received 2 March 2022
Revised 17 June 2022
Accepted 12 August 2022

KEYWORDS
Speech; vocal emotion expression; visual impairment; blindness; sociodemographic backgrounds

➤ IMPLICATIONS FOR REHABILITATION

- Technologies equipped with alternative user interfaces (e.g., Siri, Alexa, and Google Voice Assistant) meeting the needs of people with visual disabilities can promote independent living and quality of life.
- Such technologies can also be equipped with systems that can recognize emotions via users' voice, such that users can obtain services customized to fit their emotional needs or adequately address their emotional challenges (e.g., early detection of onset, provision of advice, and so on).
- The results of this study can be beneficial to health professionals (e.g., social workers) who work closely with clients who have visual disabilities (e.g., virtual telehealth sessions) as they could gain insights or learn how to recognize and understand the clients' emotional struggle by hearing their voice, which is contributing to enhancement of emotional intelligence. Thus, they can provide better services to their clients, leading to building a strong bond and trust between health professionals and clients with visual disabilities even they meet virtually (e.g., Zoom).

Introduction

Perception of emotion intensity from voice

As humans convey a great amount of information by speech signals, emotion recognition *via* auditory information has often been employed to assess one's affective states [1,2]. Plutchik [3] introduced a structural model of emotions, i.e., the shape of a cone turned upside down, to describe how various human emotions are related to each other. The cone's vertical dimension infers the intensity of emotion (e.g., *"disgust"* goes from *"dislike"* to *"loathing"*). Plutchik [3] also used colour to explain that emotions have varying degrees of intensity. For example, the darker the shade, the more intensive the emotion is. Intensity is considered as one of important cues for affective perception [4]. Holz and Larrouy-Maestri [5] empirically examined how differently people perceive the intensity of emotions when they are exposed to a sample of non-verbal vocalizations including different valence (positive and negative) emotions, ranging from low to peak intensity. They found that participants could accurately identify the intensity of the emotional expressions although other affective characteristics such as valence and emotion category were ambiguous. Banse and Scherer [6] reviewed extensively the literature and argued that people are competent at inferring emotional states by listening to others' vocal expressions.

CONTACT Hyung Nam Kim hnkim@ncat.edu North Carolina A&T State University, 1601 East Market St, Greensboro, NC 27411, USA

Emotion perception influenced by sociodemographic backgrounds

Human voice is considered as a critical communication channel that allows people to convey rich information about not only emotion but also sociodemographic backgrounds such as sex, age, social class, and geographical origin [7], which could, in turn, influence the listener's emotional perception. For example, there is evidence that different visual acuity levels are likely to affect the way people perceive emotion. A research team led by Martins [8] investigated the degree to which two groups of people with and without visual disabilities recognize emotions when exposed to non-verbal communicative emotional stimuli (e.g., tones of voice) for five different emotional states: happiness, sadness, anger, fear, and neutral. They observed that people without visual disabilities showed greater performance in all the tasks including non-emotional prosody discrimination, emotional prosody discrimination, emotional prosody naming, and conflicting emotional prosody. Bartelink et al. [9] recently conducted a systematic review on the association between unemployment and mental health. They found that unemployed people were more likely to develop mental health problems. Kessler et al. [10] observed that people who were employed again after being unemployed tended to show a significantly reduced stress level. As it is difficult for people with visual disabilities to get a job [11,12], many of them are anticipated to encounter emotional challenges and seek help. Cimarolli and Wang [13] empirically observed that employed people with visual disabilities reported a fewer number of anxiety symptoms and a greater level of life satisfaction as compared to unemployed people with visual disabilities. Age is found to be another factor influencing human emotion, leading to individual differences between younger and older people. Regardless of one's disability and ability, everyone is ageing and likely to be affected by age-related emotional challenges. Isaacowitz et al. [14] reported that younger and older people were likely to perceive different emotions when given the same emotional stimulus. Ryan et al. [15] reported that younger people outperformed the emotion recognition test when exposed to vocal expressions in a combination of vocal and facial expressions as well as in isolation. It is well documented that people with health issues are vulnerable to negative emotions. For instance, Karademas et al. [16] claimed that people who live with chronic health conditions are likely to experience a range of negative emotions (e.g., fear, anger, and sadness). Suppression of emotions (e.g., anger and hostility) is closely related with medical issues such as coronary health disease, chronic pain, rheumatoid arthritis, and cancer [17,18].

Emotion perception by people with visual disabilities

Yet, there have been a handful of publications examining the effects of auditory (speech/non-speech) stimuli on emotion perception in people with visual disabilities [8]. Dyck et al. [19] argued that children with visual disabilities are not delayed in acquiring the ability of *understanding* emotions per se; yet, they found that children with visual disabilities tended to show low performance with *recognizing* facial and vocal emotional expressions. Fryer and Freeman [20] studied the effect of verbal commentary on emotion recognition when people with visual disabilities watch movies, television, and other media forms by using Audio Description (AD). AD is a form of narration (e.g., concise and objective descriptions of actions, scenes, settings, costumes, and body languages), contributing to describing important visual details that cannot be understood from the main soundtrack alone. They found that an electronic voice (i.e., text-to-speech) delivery of such additional verbal commentary did not significantly enhance emotion elicitation as compared to a human voice delivery. They argued that human voice prosodic features (e.g., rhythm, stress, and intonation of speech) play a critical role in conveying emotions. Park and Chong [21] examined the degree to which people with visual disabilities respond to emotional stimuli in music. In response to sad music, those with visual disabilities perceived sad emotion with a significantly greater arousal and intensity levels, as compared to their peers without visual disabilities.

Emotion perception in online settings

There have been efforts to advance knowledge of vocal emotion perception in online settings. For instance, Siegert and Niebuhr [22] found that there was a significant effect of audio compression on human emotion perception, especially female voice. Perceived charisma of female speakers was significantly different from that of male speakers *via* online communication. Yamamoto and Kawahara [23] examined the degree to which Japanese children perform auditory emotion perception tasks *via* Zoom videoconferencing, e.g., the children without visual disabilities were asked to infer emotions from speech stimuli. The children performed well only when they were given the stimuli in their native language as compared to other foreign language. Koolagudi et al. [24] reported that many speech recognition systems have been developed using speech data obtained under a very well-controlled environment (e.g., speech data collected from news readers) instead of using real speech data equipped with human emotions, e.g., various intentions are incorporated to express various emotions.

Yet, little is known about how people with visual disabilities infer emotions from speech stimuli *via* information and communication technology platforms although a great number of people with disabilities take advantage of technologies to communicate with others online today. For example, a national survey [25] revealed that 71% of Americans with disabilities own a smartphone. Facebook users with visual disabilities actively engage in online social networking, which is not much different from sighted users, but also they tend to receive more feedback (i.e., comments and likes) from other Facebook users [26]. In addition, many people use online videoconferencing platforms such as Zoom to connect with others, especially amid the COVID-19 pandemic; for example, daily Zoom meeting participants increased from 10 millions in December 2019 to 200 million in March 2020 [27]. Regardless of one's ability/disability, a great amount of people use the online communication tools in that technology is becoming more accessible to those users with disabilities [28]. Zoom users with visual disabilities could, thus, go online and share a great deal of information including emotions through their voice but also perceive emotions through others' voice. However, it is unfortunate that there has still been a lack of attention paid to emotional experiences of people with visual disabilities using such information and communication technologies. Hattich and Schweizer [29] studied the degree to which people with visual disabilities are immersed and enjoy watching a movie. They found that people with visual disabilities showed poor performance in narrative understanding, and further, they considered the movie as a less thought-provoking emotional stimulus, also leading to perceiving a fewer number of positive emotions as compared to their sighted peers. Gamond et al. [30] investigated how visual disabilities would affect the specialized hemispheric lateralisation in emotion processing. Participants were given emotional stimuli (e.g., laughs and cries) *via* headphones while seating in front of a computer. In sighted participants, positive vocalizations (e.g., laughing) were detected better when presented through the right ear (i.e., largely processed in the left hemisphere), and negative vocalizations (e.g., cries) were detected better through the left ear (i.e., largely processed in the right hemisphere). However, participants with visual disabilities did not show such advantage of one ear over the other with regard to detection of emotional vocalizations.

Potential contribution of a deep understanding of human emotion perception to user-centred designs

There are numerous ways of applying the knowledge of emotional vocal expressions to system designs that can accommodate users' needs and capabilities/limitations adequately. For example, an automatic tutoring application can be designed to listen to e-learners' voice and assess their subjective status (e.g., boring, dreary, exciting, thrilling, and daunting). Such emotion-sensitive tutoring application could then accordingly change the teaching modules, e.g., the tutoring application could recommend a break, adjust the educational contents, or change the speed at which the contents are delivered to learners online [31]. The automatic emotion-sensitive systems can widely be used in various domains such as a customer service call centre [31], vocal emotion expression of a personal computer assistant [32], speech emotion recognition for e-learning [33] and emotion recognition *via* vocal social media [34]. Thus, it is essential to first obtain a deep understanding of how humans perceive emotional stimuli *via* a sense of hearing when designing a human emotion-sensitive technology.

To contribute to the advancement of fundamental knowledge of human emotion perceptions through online information and communication tools used by people with visual disabilities, this study examines the degree to which those with visual disabilities perceive emotions strongly or weakly (i.e., perceived intensity) in the context of online communication settings but also investigate the degree to which their sociodemographic backgrounds affect the emotional perceptions.

Methods

Participants

A convenience sample of 30 participants with visual disabilities contributed to this study. We found them in collaboration with community organisations such as local community centres and a local library for the blind. The organisations have access to the target population and helped us to locate participants. Participants should speak English, be 18 years of age or older, and have visual acuity worse than 20/70 [35], all of whom should not have other disabilities, severe cognitive impairment (e.g., dementia), and hearing/speaking problems. We did not ask the organisations to share their community members' contact information. The organisations were given a copy of informed consent forms and were also briefed about the study (e.g., objectives, procedures, and expected outcomes). They informed their community members of the opportunity to patriciate in this study, and we waited until they contacted us. Once we were contacted, we fully informed potential participants of this study to help them to decide whether they agree (or decline) to participate in this study. It was ensured that each participant provided us with informed consent. The detailed characteristics of participants are available in Table 1.

Table 1. Characteristics of participants.

Participants	$n = 30$
Visual acuity	
Between 20/200 and 20/400	9
Between 20/400 and 20/1200	4
Less than 20/1200, but has light perception	15
No light perception at all	2
Duration of visual disabilities (years)	17.63 ± 23.11
Age (years)	59.7 ± 17.84
Gender	
Male	11
Female	19
Race/Ethnicity	
African American	11
European American	17
Hispanic American	1
Others	1
Currently working	
Working full time	5
Working part time	2
Currently in school	2
Currently not working	
Unemployed and not looking for work	4
Unemployed and looking for work	1
Retired or disabled	16
Diagnosed with health conditions	16

Materials

Emotional speech stimuli were employed from the valid database, Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [36]. The stimuli consisted of eight emotional speeches: sad, happy, angry, calm, fearful, surprised, neutral, and disgust. They were presented using a neutral statement "*Kids are talking by the door*" and recorded in wav format (16 bit, 48 kHz).

Procedures

The participants were instructed to choose the interview time that works for their schedule and enables them to focus on the interviews. Interviews were conducted *via* Zoom. Participants' family members were not prohibited from helping the participants with operating the Zoom; however, they were asked to avoid interrupting the interviews – e.g., answering on behalf of the participants and discussing with the participants. After verbal informed consent was obtained, a demographic questionnaire was administered. A researcher read questions out loud, clearly, and slowly to help the participants understand fully. A researcher played the emotional speech files in random order. The participants were allowed to ask for replaying them as many as they wanted to hear again. Participants reported the emotion they have perceived among the following eight emotions: sad, happy, angry, calm, fearful, surprised, neutral, and disgust. They also reported the intensity level of the perceived emotion, on a rating scale from 1 (*weak*) to 8 (*strong*). This study was approved by the Institutional Review Board (IRB). Each interview lasted no longer than 60 min.

Data analysis

A descriptive statistic helped to identify an emotion with the strongest intensity among emotions participants perceived when they were exposed to each emotional stimulus. A Chi-square test was performed to investigate the degree to which the emotion with the strongest intensity is different from other emotions in terms of the perceived intensity. Mann-Whitney tests were also conducted to examine differences in the perceived intensity of emotions between participants with different sociodemographic backgrounds. Statistical analyses were performed using the IBM SPSS Statistics for Macintosh, version 24 [37].

Results

When the participants were exposed to such emotional speech stimuli as *neutral* and *angry*, they felt "neutral" and "angry" as the strongest emotion, respectively (see Table 2). However, such a congruent case was not observed when the participants were exposed to the other emotional speech. For example, when exposed to the calm, happy, fearful, and sad speech stimuli, *neutral* was the dominant emotion that the participants perceived with the strongest intensity. On the other hand, when exposed to the disgust and surprised speech stimuli, *non-neutral* emotions were strongly perceived, i.e., surprised and happy, respectively. When exposed to the *neutral* speech stimulus, the participants perceived *neutral* significantly than the other emotions (except *calm*).

We also investigated the degree to which the participants show differences in the perceived intensity of emotions depending on sociodemographic backgrounds. Mann-Whitney tests indicated that the participants with different backgrounds of health, vision, job, and age showed different perceived intensity levels of emotions (see Table 3). Healthy participants perceived angry and surprise (i.e., high-arousal emotions) with greater intensity levels as compared to their peers with health issues (e.g., chronic conditions) when exposed to the emotional speech stimuli, angry and disgust (i.e., high-arousal emotions). In contrast, those with health issues perceived neutral with a greater intensity level as compared to healthy participants when exposed to the emotional speech stimuli, fearful and sad. Participants with visual impairments perceived neutral with a greater intensity level as compared to their peers with blindness when exposed to the fearful speech stimulus. When the happy speech stimulus was given, the perceived emotion with the greatest intensity was neutral among all participants. Yet, we further investigated it by breaking the participants down into the two groups: participants *currently working* and those *currently not working*. The group of participants currently working tended to feel less neutral as compared to the group of participants currently not working. Instead, those currently working were likely to feel happy strongly (the perceived intensity Mean = 1.89 and SD = 2.85) when the happy speech stimulus was given. Both younger and older participants perceived neutral when exposed to a happy speech stimulus. Yet, older participants felt neutral much stronger that did younger participants, while younger participants felt strongly happy instead.

Discussion

This study acknowledges that there have been efforts to study human emotion expression and perception, yet there are still limitations. For example, Sen et al. [38] conducted research on age differences in vocal emotional perception, but they did not include people with visual disabilities. Further, they merely focussed on a fewer number of emotions (happy, neutral, sad, and angry) while this study covered more (sad, happy, angry,

Table 2. Wilcoxon signed rank tests between the perceived emotion with the strongest intensity level and the other perceived emotions.

Given emotional speech stimuli	Perceived emotion with the strongest intensity	The perceived emotion with the strongest intensity was compared with the other perceived emotions, in terms of the perceived intensity							
		vs. Angry	vs. Fearful	vs. Disgust	vs. Happy	vs. Sad	vs. Calm	vs. Surprise	vs. Neutral
Calm	Neutral	$z=-2.75$, $p=0.006^*$	$z=-2.49$, $p=0.013$	$z=-2.67$, $p=0.008$	$z=-1.07$, $p=0.285$	$z=-2.43$, $p=0.015$	$z=-1.32$, $p=0.188$	$z=-1.68$, $p=0.094$	–
Happy	Neutral	$z=-2.43$, $p=0.015$	$z=-2.41$, $p=0.016$	$z=-1.83$, $p=0.068$	$z=-1.65$, $p=0.1$	$z=-2.59$, $p=0.01$	$z=-.85$, $p=0.394$	$z=-.72$, $p=0.473$	–
Fearful	Neutral	$z=-2.72$, $p=0.007$	$z=-.94$, $p=0.348$	$z=-2.69$, $p=0.007$	$z=-2.09$, $p=0.037$	$z=-1.15$, $p=0.252$	$z=-1.59$, $p=0.113$	$z=-2.09$, $p=0.037$	–
Sad	Neutral	$z=-2.84$, $p=0.004^*$	$z=-1.10$, $p=0.272$	$z=-2.84$, $p=0.004^*$	$z=-2.22$, $p=0.027$	$z=-.48$, $p=0.631$	$z=-1.35$, $p=0.177$	$z=-1.87$, $p=0.061$	–
Neutral	Neutral	$z=-3.94$, $p=0.000^*$	$z=-3.94$, $p=0.000^*$	$z=-3.94$, $p=0.001^*$	$z=-2.98$, $p=0.003^*$	$z=-3.94$, $p=0.001^*$	$z=-2.51$, $p=0.012$	$z=-3.94$, $p=0.001^*$	–
Angry	Angry	–	$z=-2.28$, $p=0.023$	$z=-1.99$, $p=0.047$	$z=-3.22$, $p=0.001^*$	$z=-3.22$, $p=0.001^*$	$z=-3.22$, $p=0.001^*$	$z=-.76$, $p=0.446$	$z=-2.65$, $p=0.008$
Disgust	Surprised	$z=-2.67$, $p=0.008$	$z=-2.97$, $p=0.003^*$	$z=-.92$, $p=0.359$	$z=-1.93$, $p=0.054$	$z=-2.22$, $p=0.027$	$z=-2.22$, $p=0.027$	–	$z=-.31$, $p=0.757$
Surprised	Happy	$z=-2.69$, $p=0.007$	$z=-2.69$, $p=0.007$	$z=-2.69$, $p=0.007$	–	$z=-2.69$, $p=0.007$	$z=-.22$, $p=0.826$	$z=-1.58$, $p=0.115$	$z=-.18$, $p=0.854$

$^*\alpha \leq 0.006$ (with Bonferroni correction).

Table 3. Differences in the perceived intensity of emotions between participants with different sociodemographic backgrounds.

Given emotional speech stimuli	Intensity differences between participants with different sociodemographic backgrounds	Mann-Whitney tests
Angry	Angry perception between participants with heath issues (0.81 ± 2.29) vs. without them (5.61 ± 3.15)	$p=0.000$, $U=35.50$, $z=-3.53$
Fearful	Neutral perception between participants with visual impairment (4.00 ± 3.91) vs. blindness (1.19 ± 2.64)	$p=0.038$, $U=57.50$, $z=-2.07$
	Neutral perception between participants with health issues (3.25 ± 3.87) vs. without them (0.64 ± 1.65)	$p=0.042$, $U=72.50$, $z=-2.03$
Disgust	Surprised perception between participants working currently (5.56 ± 2.35) vs. not working currently (0.95 ± 2.40)	$p=0.000$, $U=28.00$, $z=-3.50$
	Surprised perception between participants with health issues (1.06 ± 2.29) vs. without them (3.79 ± 3.49)	$p=0.017$, $U=62.50$, $z=-2.39$
Happy	Neutral perception between participants working currently (0.00 ± 0.00) vs. not working currently (2.76 ± 3.67)	$p=0.036$, $U=58.50$, $z=-2.10$
	Neutral perception between younger (0.47 ± 1.94) vs. older participants (3.85 ± 3.81)	$p=0.005$, $U=59.00$, $z=-2.78$
Sad	Neutral perception between participants with health issues (3.63 ± 3.83) vs. without them (0.57 ± 1.51)	$p=0.018$, $U=64.50$, $z=-2.36$
Calm	Neutral perception between younger (0.76 ± 2.22) vs. older participants (3.92 ± 3.88)	$p=0.012$, $U=62.00$, $z=-2.52$

calm, fearful, surprised, neutral, and disgust). They conducted research with people living in Asia while this study did with Americans living in United States. Thus, their emotional expression and perception might be differently influenced by cultural effects [39]. In addition, they simply examined acoustic parameters (i.e., speech analysis in phonetics); on the contrary, this study was centred around human factors approach, i.e., how differently/similarly people with visual disabilities perceive various emotional stimuli and intensity. Yeshoda et al. [40] studied perception of vocal emotional prosody but included children only while this study included adults. Furthermore, they excluded those with visual disabilities but merely focussed on those with hearing disabilities. They were also limited to a few of emotions (i.e., happy, sad, and neutral) and Asians only.

When the participants in this study were exposed to the emotional speech stimuli, calm, happy, fearful, sad, and neutral, they reported that neutral was the dominant emotion they perceived with the greatest intensity. The results suggest that the perceived intensity of those stimuli (calm, happy, fearful, sad, and neutral) was not strong enough to cause the participants to feel non-neutral emotions (e.g., high- or low-arousal emotions). On the other hand, the high-arousal emotional speech stimuli angry, disgust, and surprised were able to stimulate the participants to perceive non-neutral emotions angry, surprised, and happy, respectively. Based on the results, it can be argued that the participants may have a high threshold in terms of perceiving high- or low-arousal emotions. Threshold refers to the degree to which an individual is emotionally sensitive enough to perceive a certain emotion [41]. If a speech stimulus is not intense enough to reach the threshold, it would probably fail to trigger a high/low-arousal emotion perception, resulting in neutral emotion perception. Taylor [42] claimed that negative emotional events are likely to evoke much stronger behavioural responses as compared to positive or neutral events. Cacioppo and Gardner [43] suggested that the mechanism of human emotion perception is closely related with evolutionary terms. For example, it is more critical for humans to withdraw themselves from negative stimuli/events in order to survive than to have themselves exposed to pleasant or neutral stimuli/events. It means that the withdrawal system is more responsive and primary in emotional information processing systems in humans.

Yet, the participants in this study also showed mixed patterns in that such negative stimuli as *fearful* and *sad* led to “neutral” emotion perception, while other negative stimuli *angry* and *disgust* led to “non-neutral” emotion perception. Pratto and John [44] pointed out that humans rely on automatic-evaluation processing in order to identify whether stimuli are good or bad, liked or disliked, and desirable or undesirable – i.e., a human basic and ubiquitous aspect of responding to their environment. They also discussed an automatic vigilance model of emotion in which humans allocate attentional resources to undesirable/negative emotional stimuli, leading to formation and maintenance of undesirable/negative impressions and stereotypes. Therefore, humans’ greater attention to undesirable/negative environmental cues would help them to protect themselves from any immediate harm and promote well-being. Given Pratto’s argument [44], the participants in this study may have different mechanisms of evaluating positive/negative emotional speech stimuli, leading to such mixed patterns and perceptions. For example, the participants may have considered *angry* and *disgust* emotional stimuli as emotions pertaining to more undesirable/negative emotional stimuli (leading to “non-neutral” emotion), as compared to *fearful* and *sad* emotional stimuli (leading to “neutral” emotion).

Neutral speech is typically viewed as a reference source for conversation that is not labelled with any intensity degrees [45]. Pitch, one of vocal characteristics, refers to the perceived highness or lowness of a voice [46]. As pitch can convey information about emotions, pitch helps humans to understand and interpret emotional speech adequately [47]. It is well documented that people with tone-deafness (i.e., amusic) are limited in the ability to perceive and produce pitch while communicating with others [48,49]. Amusic individuals were less likely to recognize and identify emotions in different types of vocal expressions [50]. It is, thus, hypothetically argued that the participants in this study may have different mechanisms of interpreting the vocal characteristic *pitch* of angry, disgust, fearful, and sad emotional speech stimuli, leading to individual differences in perceiving emotions. Our future research will investigate various vocal characteristics (e.g., pitch, loudness, and sound quality) [51] of emotional speech samples and their associations with the intensity of emotions perceived by people with visual disabilities.

Individual differences were also found associated with sociodemographic backgrounds such as health, vision, job, and age. The participants with health issues perceived neutral emotion with a greater intensity level when exposed to negative emotions as compared to the healthy participants. Such poor health status may have affected the way the participants understand negative emotional speech stimuli, leading to perception of neutral emotion with a greater intensity level. It is well documented that chronic disease self-efficacy contributes to reducing distress and promoting positive emotional well-being. Self-efficacy is generally defined as one’s perceived competence with coping with personal challenges (e.g., health issues) [52]. It affects one’s thinking, behavioural, and emotional experiences and outcomes. For example, patients with high self-efficacy tend to maintain positive attitudes towards life, accept their illness/challenges, and cope with difficult situations, changes, and stressors [53]. Given the literature, it can hypothetically be argued that the participants with chronic diseases who recognized *negative* speech stimuli as *neutral* emotion may have had a high-level of chronic disease self-efficacy.

Another sociodemographic factor associated with individual differences was vision. The visual acuity level of participants with blindness is poorer than that of their peers with visual impairment (including low vision). Those with visual impairment can still take advantage of their residual vision to gather contextual information from the environment, while those with blindness tend to significantly rely on a sense of hearing. Thus, those with blindness may have been more sensitive to the given emotional speech stimulus (e.g., fearful speech) and recognized it as the identical emotion (e.g., fearful emotion). From a theoretical point of view, a research team by Aron [54] introduced the highly sensitive person (HSP) theory to describe a person who is high in a personality trait known as *sensory processing sensitivity* (SPS). An individual with a high level of SPS tends to show increased emotional sensitivity and stronger reactivity to stimuli from the environment. SPS is viewed as an evolutionary beneficial survival tactic [55] as it can help people to collect resources, offer responsive care to others, and carefully observe/assess a situation to avoid threats [56]. Given our research finding and the theoretical aspects, it may be argued that those with blindness are probably more emotionally sensitive than their peers with visual impairment in response to emotional auditory stimuli. Our future research will incorporate the 27-item Highly Sensitive Person Scale (HSPS) [54] to measure the SPS level of people with visual disabilities (blindness and visual impairment) and assess the relationship between the SPS level and emotion recognition among people with visual impairment and blindness.

There is evidence that economic inactivity is associated with less positive (or negative) emotion perception [57,58]. Knabe et al. [59] found that employed people tended to perceive more positive emotions as compared to unemployed people when engaged in similar activities. As compared to employed people, their unemployed peers showed poor emotional well-being by showing less self-confident, being overwhelmed with their personal problems, and being often diagnosed with mental disorders [60]. For example, Farre et al. [60] stated that over 12% of unemployed people answered "No" to the survey question *"Have you been feeling reasonably happy, all things considered?"* while only approximately 4% of employed people answered "No" to the same question. Unemployment rates of people with visual disabilities was 4% in 1994–1995, reaching 19.8% in 2011 [11]. As many people with visual disabilities who would like to work cannot get hired, they are also anticipated to suffer from emotional challenges. Adequate emotional support and interventions are needed to help those with visual disabilities to manage their emotions.

Age-related differences were also found. Although both younger and older participants perceived neutral when exposed to a happy speech stimulus, older participants felt neutral much stronger that did younger participants. There is evidence that younger and older adults tend to perceive different emotions when exposed to the same emotional stimulus [61,62]. Riediger et al. [63] observed that older adults were less likely to perceive positive emotions when exposed to positive emotional stimuli. In their study, older adults more frequently attributed the category of "neutral" to the positive emotional stimulus, while younger counterparts more frequently attributed the category of "positive" instead. Riediger et al. [63] also found that older adults were more likely to accurately infer emotions expressed by their older peers. As the present study used emotional speech stimuli that were recorded by younger adults, future research will incorporate emotional stimuli prepared by older adults.

Limitations

This study was conducted amid the COVID-19 pandemic. The participants were probably emotionally affected by public health safety measures such as stay-at-home order and social distancing. Therefore, the results of this study may have been influenced by such external factors. In addition, the interviews were scheduled for various timeframes, e.g., ranging from morning to late afternoon as well as during weekdays and weekends. As human emotions could be affected by time [64], the results of this study may also have been affected by such diverse data collection timeframes. As qualitative interviews with participants regarding their emotional perception were not available for this study, a future research study will be adequately designed to collect them and discuss relations with emotional perception.

Conclusion

This study contributed to advancing knowledge of individual differences in perceived intensity levels of emotions that those with visual disabilities felt when exposed to various emotional speech stimuli *via* online communication (Zoom). Their perceived intensity was found to be influenced by sociodemographic backgrounds associated with health conditions, visual acuity levels, employment, and age. As many research studies on voice user interfaces tend to focus on technology advancements (e.g., accuracy of voice command recognition) over user-centred designs [65], the research findings of this study will be helpful for user experience designers in designing user-friendly voice user interfaces for users with visual disabilities. The individual differences among people with visual disabilities can be applied to designing vocalizations of voice assistants (e.g., Apple Siri and Amazon Alexa) that are emotionally understandable to those with visual disabilities. For instance, such user-centred voice user interface designs would be useful in promoting user compliance with suggestions made by computing systems (e.g., Artificial Intelligence) for nutrition coaching, students' guides, emergency responses, and so on [66]. According to the *7-38-55* rule by Mehrabian and Ferris [67], only 7% of message pertaining to feelings and attitudes is conveyed through verbal (spoken words), while 38% and 55% are conveyed through vocal (voice characteristics such as volume, tone, and pitch) and facial expressions, respectively. As those with visual disabilities are less likely to rely on facial expressions to perceive emotions, user-friendly voice user interface designs are expected to significantly contribute to advancements in human-computer interactions for users with visual disabilities. The results of this study can serve as fundamental knowledge foundations that will be helpful for many other stakeholders. For example, engineers can use the results to design emotion-sensitive technology applications that are programmed to assess emotions of users with visual disabilities *via* voice user interfaces and offer resources customized to help those users to regulate their emotional challenges. Social workers and health professionals can also refer to the results of this study to provide better services to their clients with visual disabilities by being emotionally connected with them (e.g., high empathy during verbal communication).

Future research will investigate the degree to which people with visual disabilities perceive different intensity levels of emotions depending on various conditions associated with a chronic disease self-efficacy level, a sensory processing sensitivity, vocal characteristics of emotional speech samples, and emotional speech samples recorded by people with different sociodemographic backgrounds such as age, race, and ethnicity. There are a range of speech analysis technologies including WaveSurfer [68], Speech Analyser [69], Speech Filing System [70], Praat [71], and Sound Tools eXtended [72]. Those technologies may be incorporated in future research in order to obtain a deeper understanding of a range of difference speech samples. For example, the technologies could help to extract vocal characteristics (e.g., pitch, tone, and volume) from emotional speech stimuli and then explore whether there are relationships between those characteristics and the perceived emotions of people with visual disabilities. Future research will also consider conducting the same procedure by inviting people without visual disabilities so that the result could be compared with that of their peers with visual disabilities. The advance understanding of individual differences between the two groups would contribute to developing inclusive voice user interface designs for all including people with visual disabilities.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This material is based upon work supported by the National Science Foundation of the United States under Grant No. 1831969.

ORCID

Hyung Nam Kim http://orcid.org/0000-0003-1443-2122

References

[1] Ooi CS, Seng KP, Ang L-M, et al. A new approach of audio emotion recognition. Expert Syst Appl. 2014;41(13): 5858–5869.
[2] Pell MD, Skorup V. Implicit processing of emotional prosody in a foreign versus native language. Speech Commun. 2008;50(6):519–530.
[3] Plutchik R. The psychology and biology of emotion. New York: HarperCollins College Publishers; 1994.
[4] Schmidt J, Janse E, Scharenborg O. Perception of emotion in conversational speech by younger and older listeners. Front Psychol. 2016;7:781.
[5] Holz N, Larrouy-Maestri P, Poeppel D. The paradoxical role of emotional intensity in the perception of vocal affect. Sci Rep. 2021;11(1):1–10.
[6] Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 1996;70(3):614–636.
[7] Karpf A. The human voice: How this extraordinary instrument reveals essential clues about who we are. New York: Bloomsbury Publishing USA; 2006.
[8] Martins AT, Faísca L, Vieira H, et al. Emotional recognition and empathy both in deaf and blind adults. J Deaf Stud Deaf Educ. 2019;24(2):119–127.
[9] Bartelink VH, Zay Ya K, Guldbrandsson K, et al. Unemployment among young people and mental health: a systematic review. Scand J Public Health. 2020;48(5): 544–558.
[10] Kessler RC, Turner JB, House JS. Unemployment, reemployment, and emotional functioning in a community sample. Am Sociolog Rev. 1989;54(4):648–657.
[11] McDonnall MC, Sui Z. Employment and unemployment rates of people who are blind or visually impaired: estimates from multiple sources. J Visual Impairment Blindness. 2019;113(6):481–492.
[12] American Foundation for the Blind. *Reviewing the disability employment research on people who are blind or visually impaired: Key takeaways*. 2020. Available from: https://www.afb.org/research-and-initiatives/employment/reviewing-disability-employment-research-people-blind-visuallys.
[13] Cimarolli V, Wang S. Differences in social support among employed and unemployed adults who are visually impaired. J Visual Impair Blindn. 2006;100(9):545–556.
[14] Isaacowitz DM, Löckenhoff CE, Lane RD, et al. Age differences in recognition of emotion in lexical stimuli and facial expressions. Psychol Aging. 2007;22(1):147–159.
[15] Ryan M, Murray J, Ruffman T. Aging and the perception of emotion: processing vocal expressions alone and with faces. Exp Aging Res. 2010;36(1):1–22.
[16] Karademas EC, Tsalikou C, Tallarou M-C. The impact of emotion regulation and illness-focused coping strategies on the relation of illness-related negative emotions to subjective health. J Health Psychol. 2011;16(3):510–519.
[17] Nyklicek I, Vingerhoets A, Denollet J. Emotional (non-)expression and health: data, questions, and challenges. Psychol Health. 2002;17(5):517–528.
[18] Panagopoulou E, Kersbergen B, Maes S. The effects of emotional (non-) expression in (chronic) disease: a meta-analytic review. Psychol Health. 2002;17(5):529–545.
[19] Dyck MJ, Farrugia C, Shochet IM, et al. Emotion recognition/understanding ability in hearing or vision-impaired children: do sounds, sights, or words make the difference? J Child Psychol Psychiatry. 2004;45(4):789–800.
[20] Fryer L, Freeman J. Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In: Felnhofer A. and Kothgassner OD, editors. *Challenging presence: proceedings of the 15th international conference on presence*; 2014. Facultas. p. 99–107.
[21] Park HY, Chong HJ. A comparative study of the perception of music emotion between adults with and without visual impairment. Psychol Music. 2019;47(2):225–240.
[22] Siegert I, Niebuhr O. Case report: women, be aware that your vocal charisma can dwindle in remote meetings. Front Commun. 2021;5:611555.
[23] Yamamoto HW, Kawahara M, Tanaka A. A web-based auditory and visual emotion perception task experiment with children and a comparison of lab data and web data. Front Psychol. 2021;12:3388.
[24] Koolagudi SG, Maity S, Kumar VA, et al. IITKGP-SESC: speech database for emotion analysis. In: Ranka S, editor. Contemporary computing. IC3 2009. Communications in computer and information science. Springer: Berlin, Heidelberg; 2009. p. 485–492.
[25] Morris JT, Jones ML, Sweatman WM. Wireless technology use by people with disabilities: a national survey. J Technol Persons Disabil. 2016;1:101–113.
[26] Wu S, Adamic LA. Visually impaired users on an online social network. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Toronto, ON, Canada; 2014.
[27] Chawla A. Coronavirus (COVID-19)–'Zoom'application boon or bane. SSRN. 2020;2020:1–10.
[28] Dolamore S. Accessibility features in ZOOM to improve equity in the MPA classroom. J Public Affairs Educ. 2021; 27(3):376–374.
[29] Hättich A, Schweizer M. I hear what you see: effects of audio description used in a cinema on immersion and enjoyment in blind and visually impaired people. Br J Visual Impair. 2020;38(3):284–298.
[30] Gamond L, Vecchi T, Ferrari C, et al. Emotion processing in early blind and sighted individuals. Neuropsychology. 2017; 31(5):516–524.
[31] Fragopanagos N, Taylor JG. Emotion recognition in human–computer interaction. Neural Netw. 2005;18(4): 389–405.
[32] Kessens JM, Neerincx MA, Looije R, et al. Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children, in 2009 3rd international conference on affective computing and intelligent interaction and workshops. Amsterdam, The Netherlands: IEEE; 2009. p. 1–7.
[33] Bahreini K, Nadolski R, Westera W. Towards real-time speech emotion recognition for affective e-learning. Educ Inf Technol. 2016;21(5):1367–1386.
[34] Dai W, Han D, Dai Y, et al. Emotion recognition and affective computing on vocal social media. Inform Manag. 2015; 52(7):777–788.
[35] World Health Organization. *Change the definition of blindness.* 2008. Available from: https://www.who.int/health-topics/blindness-and-vision-loss.

[36] Livingstone SR, Russo FA. The ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PLoS One. 2018;13(5):e0196391.

[37] IBM Corp. IBM SPSS statistics for Macintosh, version 24.0.Armonk, NY: IBM Corp; 2016.

[38] Sen A, Isaacowitz D, Schirmer A. Age differences in vocal emotion perception: on the role of speaker age and listener sex. Cogn Emot. 2018;32(6):1189–1204.

[39] Engelmann J, Pogosyan M. Emotion perception across cultures: the role of cognitive mechanisms. Front Psychol. 2013;4:118.

[40] Yeshoda K, Raveendran R, Konadath S. Perception of vocal emotional prosody in children with hearing impairment. Int J Pediatr Otorhinolaryngol. 2020;137:110252.

[41] Chaturvedi M, Chander R. Development of emotional stability scale. Ind Psychiatry J. 2010;19(1):37–40.

[42] Taylor SE. Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. Psychol Bull. 1991;110(1):67–85.

[43] Cacioppo JT, Gardner WL. Emotion. Annu Rev Psychol. 1999;50(1):191–214.

[44] Pratto F, John OP. Automatic vigilance: the attention-grabbing power of negative social information. J Pers Soc Psychol. 1991;61(3):380–391.

[45] Tao J, Kang Y, Li A. Prosody conversion from neutral speech to emotional speech. IEEE trans Audio Speech Lang Pro. 2006;14(4):1145–1154.

[46] Klofstad CA, Nowicki S, Anderson RC. How voice pitch influences our choice of leaders: when candidates speak, their vocal characteristics–as well as their words–influence voters' attitudes toward them. Am Sci. 2016;104(5):282–288.

[47] Lolli S, Lewenstein AD, Basurto J, et al. Sound frequency affects speech emotion perception: results from congenital Amusia. Front Psychol. 2015;6:1340.

[48] Hyde KL, Peretz I. Brains that are out of tune but in time. Psychol Sci. 2004;15(5):356–360.

[49] Vuvan DT, Nunes-Silva M, Peretz I. Meta-analytic evidence for the non-modularity of pitch processing in congenital Amusia. Cortex. 2015;69:186–200.

[50] Lima CF, Brancatisano O, Fancourt A, et al. Impaired socio-emotional processing in a developmental music disorder. Sci Rep. 2016;6(1):1–13.

[51] Maryn Y. Practical acoustics in clinical voice assessment: a Praat primer. Perspect ASHA SIGs. 2017;2(3):14–32.

[52] Mineva K, Peeva K. Predictors of emotional illness perception in patients with chronic diseases. Trakia J Sci. 2016;14(4):367–372.

[53] Ebrahimi Belil F, Alhani F, Ebadi A, et al. Self-efficacy of people with chronic conditions: a qualitative directed content analysis. J Clin Med. 2018;7(11):411.

[54] Aron EN, Aron A. Sensory-processing sensitivity and its relation to introversion and emotionality. J Pers Soc Psychol. 1997;73(2):345–368.

[55] Bolders AC, Tops M, Band GP, et al. Perceptual sensitivity and response to strong stimuli are related. Front Psychol. 2017;8:1642.

[56] Acevedo B, Aron E, Pospos S, et al. The functional highly sensitive brain: a review of the brain circuits underlying sensory processing sensitivity and seemingly related disorders. Phil Trans R Soc B. 2018;373(1744):20170161.

[57] Batic-Mujanovic O, Poric S, Pranjic N, et al. Influence of unemployment on mental health of the working age population. Mater Sociomed. 2017;29(2):92–96.

[58] Llena-Nozal A, OECD F. *Are all jobs good for your health? The impact of work status and working conditions on mental health*. 2008. Available from: https://www.oecd.org/els/emp/oecdemploymentoutlook2008.htm#CH4.

[59] Knabe A, Rätzel S, Schöb R, et al. Dissatisfied with life but having a good day: time-use and well-being of the unemployed. The Economic Journal. 2010;120(547):867–889.

[60] Farré L, Fasani F, Mueller H. Feeling useless: the effect of unemployment on mental health in the great recession. IZA J Labor Econ. 2018;7(1):1–34.

[61] Ebner NC, Fischer H. Emotion and aging: evidence from brain and behavior. Front Psychol. 2014;5:996.

[62] Cortes DS, Tornberg C, Bänziger T, et al. Effects of aging on emotion recognition from dynamic multimodal expressions and vocalizations. Sci Rep. 2021;11(1):1–12.

[63] Riediger M, Studtmann M, Westphal A, et al. No smile like another: adult age differences in identifying emotions that accompany smiles. Front Psychol. 2014;5:480.

[64] English T, Carstensen LL. Emotional experience in the mornings and the evenings: consideration of age differences in specific emotions by time of day. Front Psychol. 2014;5:185.

[65] Namkung K. Research on emotional factors and voice trend by country to be considered in designing AI's Voice-An analysis of interview with experts in Finland and Norway. J Korea Convergence Soc. 2020;11(9):91–97.

[66] Chidambaram V, Chiang Y-H, Mutlu B. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. Boston, Massachusetts; 2012. p. 293–300.

[67] Mehrabian A, Ferris SR. Inference of attitudes from nonverbal communication in two channels. J Consult Psychol. 1967;31(3):248–252.

[68] Sjölander K, Beskow J. Wavesurfer-an open source speech tool, in sixth international conference on spoken language processing (ICSLP). Beijing, China; 2000. p. 464–467.

[69] SIL Language Technology, *Speech Analyzer*. 2022. https://software.sil.org/speech-analyzer/.

[70] Huckvale M. *Speech Filing System*. 2022. https://www.phon.ucl.ac.uk/resource/sfs/.

[71] Boersma P, Van Heuven V. Speak and unSpeak with PRAAT. Glot International. 2001;5(9/10):341–347.

[72] Noll A, Stuefer J, Klingler N, et al. Sound Tools eXtended (STx) 5.0-A powerful sound analysis tool optimized for speech. In INTERSPEECH; 2019.