



ORIGINAL RESEARCH

Voice emotion recognition by Mandarin-speaking pediatric cochlear implant users in Taiwan

Yung-Song Lin MD^{1,2}  | Che-Ming Wu MD^{3,4,5} | Charles J. Limb MD⁶  |
Hui-Ping Lu MS⁷ | I. Jung Feng PhD⁸ | Shu-Chen Peng PhD⁹ |
Mickael L. Deroche PhD¹⁰ | Monita Chatterjee PhD¹¹

¹Department of Otolaryngology, Chi Mei Medical Center, Tainan, Taiwan

²Department of Otolaryngology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

³Department of Otorhinolaryngology, New Taipei Municipal TuCheng Hospital (built and operated by Chang Gung Medical Foundation), New Taipei City, Taiwan

⁴Department of Otorhinolaryngology, Chang Gung Memorial Hospital, Taoyuan, Taiwan

⁵School of Medicine, Chang Gung University, Taoyuan, Taiwan

⁶School of Medicine, University of California San Francisco, San Francisco, California, USA

⁷Center of Speech and Hearing, Department of Otolaryngology, Chi Mei Medical Center, Tainan, Taiwan

⁸Institute of Precision Medicine, National Sun Yat-sen University, Kaohsiung, Taiwan

⁹Center for Devices and Radiological Health, United States Food and Drug Administration, Silver Spring, Maryland, USA

¹⁰Concordia University, Montréal, Québec, Canada

¹¹Boys Town National Research Hospital, Omaha, Nebraska, USA

Correspondence

Yung-Song Lin, Department of Otolaryngology, Chi Mei Medical Center, 901 Chung Hua Road, Yung-Kan District, Tainan 71004, Taiwan.
Email: kingear@gmail.com

Abstract

Objectives: To explore the effects of obligatory lexical tone learning on speech emotion recognition and the cross-culture differences between United States and Taiwan for speech emotion understanding in children with cochlear implant.

Methods: This cohort study enrolled 60 cochlear-implanted (cCI) Mandarin-speaking, school-aged children who underwent cochlear implantation before 5 years of age and 53 normal-hearing children (cNH) in Taiwan. The emotion recognition and the sensitivity of fundamental frequency (F_0) changes for those school-aged cNH and cCI (6–17 years old) were examined in a tertiary referred center.

Results: The mean emotion recognition score of the cNH group was significantly better than the cCI. Female speakers' vocal emotions are more easily to be recognized than male speakers' emotion. There was a significant effect of age at test on voice recognition performance. The average score of cCI with full-spectrum speech was close to the average score of cNH with eight-channel narrowband vocoder speech. The average performance of voice emotion recognition across speakers for cCI could be predicted by their sensitivity to changes in F_0 .

Conclusions: Better pitch discrimination ability comes with better voice emotion recognition for Mandarin-speaking cCI. Besides the F_0 cues, cCI are likely to adapt their voice emotion recognition by relying more on secondary cues such as intensity and duration. Although cross-culture differences exist for the acoustic features of voice emotion, Mandarin-speaking cCI and their English-speaking cCI peer expressed a positive effect for age at test on emotion recognition, suggesting the learning effect and brain plasticity. Therefore, further device/processor development to improve presentation of pitch information and more rehabilitative efforts are needed to improve the transmission and perception of voice emotion in Mandarin.

Level of evidence: 3.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Laryngoscope Investigative Otolaryngology* published by Wiley Periodicals LLC on behalf of The Triological Society.

Funding information

National Institutes of Health, NIH,
R01-DC014233-01.

KEYWORDS

cochlear implant, lexical tone, pitch discrimination, voice emotion

1 | INTRODUCTION

Vocal expression of emotion is necessary for social interaction across civilizations.¹⁻³ In addition to facial expressions, people often use prosodic vocal cues to show their emotions.⁴ Consequently, when prosodic vocal cues are absent, the expression and perception of emotion are hindered, negatively affecting social interactions and development. These have been frequently observed in children with cochlear implants (CIs).⁵⁻⁷

Although cochlear implant (CI) development has had remarkable achievements allowing people with profound hearing loss to hear lexical meaning in certain environments,⁸⁻¹⁰ limitations remain for present CI systems. Limitations in a CI system to transmit prosodic cues (e.g., pitch) result in limitations in the user's interpretation and communication of voice emotion.¹¹ For tonal languages, slow-pitch changes convey prosodic/emotional information, whereas rapid inflections within syllables convey meaning.^{12,13} The unique demands for adequate pitch perception in tonal language may alter the fundamental frequency (F0)-processing mechanisms of a developing auditory system. This might be altered further in children with CIs. The present study supposes that (secondary) covarying cues such as changes in intensity and duration convey the same information to some extent when F0 processing is degraded.¹⁴ To further explore the effects of obligatory lexical tone learning on speech emotion recognition, the current study was conducted on Mandarin-speaking children in Taiwan. This work, involving collaborations with labs in the United States and Taiwan, is expected to explore the cross-culture differences for speech emotion understanding¹⁵ and reveal important implications for both CI technology and rehabilitative therapies for children with CIs.

2 | MATERIALS AND METHODS

2.1 | Study design and oversight

This study was part of one multicenter, retrospective cohort research, sponsored by National Health Institute (NIH R01-DC014233-01). English-speaking subjects were recruited and tested at the Johns Hopkins University School of Medicine in Baltimore, MD, and Boys Town National Research Hospital in Omaha, NE. Results and methods of the study for English-speaking participants have been published in a prior study.¹⁶ Mandarin-speaking children were recruited and tested at Chi Mei Medical Center and Chang Gung Memorial Hospital in Taiwan. Sixty cochlear-implanted (cCI) and 53 normal-hearing (cNH) Mandarin-speaking children participated in this study (Tables 1 and 2). The Institutional Review Board and the Committee of Human Subjects Protection of Chi Mei Medical

Center approved this study. The author(s) declare(s) that there is no conflict of interest.

In this study, we measured the emotion recognition by school-aged cNH and cCI (6–17 years old). The cNH performed the task with both original (full-spectrum) speech and spectrally degraded, 4-, 8-, and 16-channel narrowband vocoder (NBV) speech. As it was expected that the cCI would have difficulty in the task, the stimuli were recorded in a child-directed manner.¹⁷ The sensitivity of F0 changes was also tested for the cNH and cCI. All participants gave informed consent prior to participation.

2.2 | Participants

Sixty profoundly hearing-impaired children without physical and visual deficit (disabilities) who underwent cochlear implantation before 5 years of age (32 boys, 28 girls, age range: 6.41–17.38 years, mean age 10.23 ± 3.22 years) and 53 cNH (21 boys, 32 girls, age range: 6.52–16.78 years, mean age 10.96 ± 2.92 years) participated in this study (Tables 1 and 2). There was no significant mean age difference between the two groups ($t = -1.193$, $p = .236$). Table 1 showed the clinical characteristics of the participating children with CI. For the cCI, 51 were implanted with Nucleus, 5 were implanted with MED-EL devices, and 4 were implanted with AB devices. Test of Nonverbal Intelligence—Third edition was used to evaluate the general intelligence of the participants. There was no significant mean intelligence difference between the two groups (cCI: mean = 102.91, SD = 12.08; cNH: mean = 109.38, SD = 11.40; $p = .09$).

2.3 | Tasks

2.3.1 | Recording

Twelve emotionally neutral sentences (Table 3) from the Hearing In Noise Test (HINT) corpus were translated from English to Mandarin and recorded by two speakers (one male and one female) in five different emotions (happy, scared, neutral, sad, and angry) in a child-directed manner. The two speakers were 25 and 27 years old and native speakers of Mandarin.

2.3.2 | Listening task

Inclusion criteria for the participants were (1) children aged from 6 to 18 with normal hearing, (2) prelingually deaf children aged from 6 to 18,

TABLE 1 Participants, children with cochlear implants

Participants	Sex	Age at implantation	Age at test	Device experience	Average residual hearing	TONI-3	Device	Insertion length/active channels	Strategy	Daily listening condition
CM13	F	2.71	7.50	4.80	90	79	Nucleus 24RE	Full/20	ACE	CI only
CX15	M	2.94	7.92	4.98	100	124	Nucleus 24RE	Full/20	ACE	CI only
CM16	F	3.21	9.47	6.26	>100	125	Nucleus 24RE	Full/20	ACE	CI only
CX17	F	2.73	7.18	4.45	90	112	Nucleus 24RE	Full/20	ACE	CI only
CG18	M	1.85	15.81	13.97	90	111	Nucleus 24RE	Full/21	ACE	CI only
NT40	M	2.27	12.55	10.28	100	111	MED-EL Pulsar	Full/12	FSP	CI only
ZX43	M	2.16	8.14	5.98	100	97	Nucleus 24RE	Full/21	ACE	CI only
ZX44	M	2.40	6.81	4.41	100	126	AB HiRes90k	Full/120	HiRes-P/Fidelity-120	CI only
MJ45	M	4.52	8.23	3.71	85	95	Nucleus 24RE	Full/20	ACE	CI only
CG12	M	2.54	12.76	10.22	100	105	Nucleus 24CS	Full/20	ACE	CI only
CX19	M	3.80	15.50	11.70	90	98	Nucleus 24CS	Full/20	ACE	CI only
CX20	F	1.56	8.95	7.39	100	105	Nucleus 24RE	Full/21	ACE	CI only
CG21	M	2.54	7.04	4.50	90	92	Nucleus 24RE	Full/20	ACE	CI only
ZX22	M	2.42	7.95	5.53	>100	93	AB HiRes90k	Full/120	HiRes-P/Fidelity-120	CI only
CG23	F	1.50	7.28	5.78	90	106	Nucleus 24RE	Full/21	ACE	CI only
CR24	M	1.92	7.83	5.91	100	122	Nucleus 24RE	Full/20	ACE	CI only
CG25	M	2.52	8.57	6.05	90	127	Nucleus 24RE	Full/21	ACE	CI only
CX26	F	2.18	10.35	8.18	100	83	Nucleus 24RE	Full/21	ACE	CI only
XG27	M	2.13	15.94	13.81	>100	81	Nucleus 24RE	Full/21	ACE	CI only
CG28	F	2.74	8.77	6.03	100	95	Nucleus 24RE	Full/20	ACE	CI only
CX29	M	1.54	7.24	5.70	90	112	Nucleus 24RE	Full/21	ACE	CI only
CL30	F	2.03	17.21	15.17	100	Over	Nucleus 24RE	Full/20	ACE	CI only
CG31	M	2.81	15.76	12.95	100	94	Nucleus 24CS	Full/20	ACE	CI only
MC37	M	1.29	9.56	8.27	90	126	Nucleus 24RE	Full/21	ACE	CI only
XG32	F	2.33	9.65	7.32	100	94	Nucleus 24RE	Full/20	ACE	CI only
XG33	F	3.31	14.57	11.27	90	109	Nucleus 24CS	Full/20	ACE	CI only
CG34	M	1.60	9.78	8.18	90	116	Nucleus 24RE	Full/21	ACE	CI only
CM35	F	2.30	13.56	11.25	>100	86	Nucleus24CS	Full/20	N24	CI only
CM51	F	1.96	6.59	4.63	100	114	MED-EL Sonata	Full/12	FSP	CI only
CG36	F	2.60	9.03	6.43	90	105	Nucleus 24RE	Full/20	ACE	CI only
CG37	F	3.06	10.10	7.05	90	100	Nucleus 24RE	Full/20	ACE	CI only
CG08	F	2.53	11.13	8.60	90	116	Nucleus 24RE	Full/20	ACE	CI only
CR30	F	3.29	17.38	14.09	100	Over	Nucleus 24RE	Full/20	ACE	CI only
XG38	M	1.62	16.69	15.06	90	Over	Nucleus 24RE	Full/21	ACE	CI only
XG39	M	3.32	7.87	4.55	>100	103	Nucleus 24RE	Full/19	ACE	CI only
CG40	M	2.64	10.13	7.49	90	115	Nucleus 24RE	Full/20	ACE	CI only
XG04	F	3.77	15.27	11.50	90	93	Nucleus 24SC	Full/20	ACE	CI only
CM49	M	1.40	9.75	8.35	100	109	Nucleus 24RE	Full/21	ACE	CI only
CG41	F	1.93	7.15	5.22	90	110	Nucleus 24RE	Full/21	ACE	CI only
CX42	F	2.44	7.80	5.36	90	112	Nucleus 24RE	Full/20	ACE	CI only
CX43	M	4.44	11.30	6.86	90	97	Nucleus 24RE	Full/20	ACE	CI only
CG44	F	3.60	9.21	5.61	90	87	Nucleus 24RE	Full/19	ACE	CI only
CL45	M	1.01	7.29	6.28	100	97	Nucleus 24RE	Full/21	ACE	CI only
MG12	F	2.63	15.84	13.22	>100	98	Nucleus 24RE	Full/21	ACE	CI only
MG46	F	1.07	9.06	7.98	100	115	Nucleus 24RE	Full/21	ACE	CI only
CG47	M	1.11	6.66	5.55	100	122	Nucleus 24RE	Full/21	ACE	CI only

TABLE 1 (Continued)

Participants	Sex	Age at implantation	Age at test	Device experience	Average residual hearing	TONI-3	Device	Insertion length/active channels	Strategy	Daily listening condition
XG48	M	2.87	17.19	14.32	90	Over	Nucleus 24RE	Full/20	ACE	CI only
CG49	F	2.84	7.87	5.03	90	101	Nucleus 24RE	Full/20	ACE	CI only
CG51	M	1.52	7.69	6.17	100	97	Nucleus 24RE	Full/21	ACE	CI only
CM53	F	1.28	10.52	9.23	100	122	MED-EL Concerto	Full/12	FSP	CI only
CM54	M	1.13	7.30	6.17	90	106	AB HiRes90k	Full/120	HiRes-P/Fidelity-120	CI only
CM52	F	1.64	13.33	11.70	100	103	Nucleus 24RE	Full/20	ACE	CI only
XG53	M	2.36	9.50	7.14	90	103	Nucleus 24RE	Full/20	ACE	CI only
CG54	M	4.21	9.50	5.29	90	92	Nucleus 24RE	Full/20	ACE	CI only
CM55	M	1.15	8.31	7.16	100	106	Nucleus 24RE	Full/21	ACE	CI only
CM56	F	1.96	6.41	4.44	100	118	MED-EL Concerto	Full/12	FSP	CI only
CM57	F	4.42	6.46	2.04	90	108	MED-EL Concerto	Full/12	FSP	CI only
XG56	M	3.68	7.24	3.56	90	103	AB HiRes90k	Full/120	HiRes-P/Fidelity-120	CI only
CG57	F	3.22	13.43	10.21	90	95	Nucleus 24RE	Full/20	ACE	CI only
XG58	M	3.07	8.97	5.89	90	95	Nucleus 24RE	Full/19	ACE	CI only

Abbreviations: CI, Cochlear implant; F, female; M, male; TONI-3, Test of Nonverbal Intelligence—Third edition; ACE, advanced combination encoder; FSP, fine structure processing.

who underwent cochlear implantation at <5 years of age, and (3) all participants should not have any other physical and visual disability. All children received a hearing test (pure-tone audiometry and sound field audiometry test) and a nonverbal intelligence test before starting the assigned task. The mother's educational level, an important predictor of performance,^{18,19} was also recorded. For the participating children, the nonverbal intelligence was measured using the matrix reasoning and block design subtests of the Wechsler Abbreviated Scale of Intelligence²⁰; linguistic ability was measured using Peabody Picture Vocabulary Test.²¹

2.4 | Stimuli

2.4.1 | Emotion recognition

Speakers for the recording task were seated in a sound-treated booth, positioned 12 in. in front of a SHURE SM63 microphone with Marantz PMD661 solid-state recorder, and produced the sentences in the five emotions three times each. The sentences selected from the HINT corpus were translated from English to Mandarin based on their semantically emotion-neutral content. Using Adobe Audition version 1.5 software, the original recorded audio files (44.1 kHz sampling rate, 16 bit) were edited. Noise-vocoded versions of these sentences were also created in 4, 8, and 16 channels using AngelSim software (Emily Shannon Fu Foundation, www.tigerspeech.com). The method for noise vocoding paralleled as described by Shannon et al.²² All stimuli were presented via a soundcard, and a single loudspeaker was located approximately 2 ft from the listeners, at an average level of 65 dB sound pressure level (SPL).

2.4.2 | Acoustic analysis of the stimuli sentences

Praat v. 5.3.56 was used to analyze the range of intensity (max – min in dB), mean intensity (dB SPL), overall duration (s), mean *F0* height (Hz), and *F0* range (ratio of maximum to minimum *F0*) across all recordings (Boersma 2001).²³ Repeated measures analyses of variance were applied for the results of acoustic analysis. Discriminability of the stimuli for different pairs of emotions was further analyzed. All discriminabilities (*d'*) within the matrix for each cue were summed to be a measure of the net discriminability provided by that cue. Figure 1 revealed the acoustic features of all sentences. Figure 1 shows that male speakers' *F0* height and intensity cues carried the greater weight of discriminability. In contrast, the female speaker's voice did not emphasize specific acoustic cues as discriminability was more homogeneously spread across the five metrics (even though *F0* height and mean intensity were again the most useful cues). By comparison, we also plotted the analyses of our previous study in English-speaking cCI in the bottom right panel of Figure 1.¹⁶ The discriminability measure (*d'*) was formulated as described by Charterjee et al.¹⁶

2.4.3 | Dynamic *F0* changes

F0-sweep stimuli were generated from broadband harmonic complexes with 100 partials, all in sine phase with equal amplitude (sampling rate of 44.1 kHz). The overall signal was low-pass-filtered at 10 kHz to ensure similar access to the bandwidth by cCI and cNH listeners. All stimuli were 300 ms long with 30-ms onset and offset ramps. The *F0* of the complex varied linearly from beginning to end with 12 final/initial *F0* ratios (sweep rates of 0.5, 1, 2, 4, 8, and

TABLE 2 Participants, normal-hearing children

Participants	Gender	Age at testing	TONI-3 score	Average PTA
1	F	10.44	102	5.83
2	M	13.78	116	7.50
3	M	12.71	113	10.00
4	M	10.44	116	8.33
5	F	10.76	94	5.83
6	F	6.92	105	10.00
7	F	13.42	89	2.08
8	F	13.54	100	2.92
9	F	10.71	100	14.58
10	F	16.22	122	2.92
11	M	8.11	95	10.42
12	M	10.02	111	7.08
13	F	7.04	119	5.00
14	F	16.43	94	5.00
15	F	9.23	118	3.75
16	F	11.09	92	10.42
17	F	14.01	107	3.75
18	F	16.78	--	5.00
19	F	9.02	113	5.00
20	F	7.76	120	6.25
21	F	7.38	121	9.17
22	F	15.58	108	8.33
23	M	13.76	91	10.42
24	F	13.16	111	4.58
25	F	13.76	100	8.33
26	M	8.72	113	7.92
27	F	7.92	124	7.92
28	F	9.81	118	6.67
29	F	15.27	140	9.17
30	F	10.84	109	9.58
31	M	14.51	94	9.58
32	M	9.10	107	7.92
33	M	10.95	104	11.25
34	M	9.91	113	7.08
35	M	10.09	104	5.00
36	F	9.25	92	6.25
37	M	8.62	120	11.25
38	F	12.10	106	9.17
39	F	14.03	129	10.83
40	F	15.33	112	3.33
41	F	15.15	109	3.75
42	F	14.69	124	0.00
43	M	9.02	107	5.83
44	F	8.52	105	6.25
45	M	10.30	129	5.83
46	F	7.39	102	12.08

TABLE 2 (Continued)

Participants	Gender	Age at testing	TONI-3 score	Average PTA
47	M	9.29	115	10.00
48	M	8.89	110	7.50
49	M	8.45	110	8.75
50	M	8.21	91	10.42
51	M	7.25	106	7.92
52	M	6.52	108	2.92
53	F	8.44	130	5.42
Average		10.96	109.38	7.24

Abbreviations: F, Female; M, male; PTA, Pure-tone audiometry; TONI-3, Test of Nonverbal Intelligence—Third edition.

16 semitones per second), yielding final/initial F_0 ratios ranging from 0.25 semitones (an increase of 1.4% over initial F_0) to 8 semitones (an increase of 58.74% over initial F_0). The starting F_0 was chosen randomly from one trial to the next from one of 10 bins uniformly distributed between 120 and 140 Hz, without replacement. In the discrimination task, the stimuli with opposite sweep directions had the same F_0 range. All stimuli were equalized at 65 dB SPL and presented with a ± 3 dB level roving. There were six rates of F_0 sweep and two directions (rising or falling). For both tasks, each experimental condition was repeated 10 times (120 trials).

2.5 | Test procedures

2.5.1 | Emotion recognition test

The participants heard each sentence and indicated which emotion was best associated with it by clicking on one of the five choices on the screen. The 12 sentences and 5 emotions were fully randomized within each condition. Four conditions were available for testing in all: full-spectrum speech, 16-channel NV speech, 8-channel NV speech, and 4-channel NV speech. All cCI heard only full-spectrum speech, and cNH heard all four conditions (randomized order). Sentences were presented in blocks of a given speaker (male or female, also counterbalanced) and condition. Listeners were given passive training with sentences not used in testing to familiarize themselves with the speakers' styles. Participants were encouraged to take breaks between blocks. No feedback was provided during the test.

2.5.2 | Discrimination of F_0 changes

Participants completed 20 practice trials, using the highest sweep rate with rising and falling directions, and no level-roving. The tasks used a child-friendly interface with an animated cartoon figure of an animal of their choice: a smiley face providing encouragement for correct response and a sad face for incorrect response in the Task. Points were earned after completing certain numbers of trials to keep the child engaged. The task (discrimination) used a three-interval, two-alternative

forced-choice procedure, presenting a reference *F0*-sweep stimulus, with either a rising or falling tone. The other stimuli were presented, one identical to the reference and the other with opposite direction (the latter two in random order). The listener was asked which, of Intervals 2 and 3, sounded different from the reference (Interval 1). Reaction

times were recorded for each trial. Percentages of correct scores were finally converted into d' and β values for statistical analyses.

3 | RESULTS

3.1 | Error patterns of emotion recognition

Figure 2 shows the error patterns for the cCI and cNH groups of listeners, male and female sentences, and under each condition of spectral resolution tested. The cells are color-coded to represent the strength of the numerical values, whereas the actual values are also indicated. The matrix patterns for cNH become increasingly diagonally dominant as spectral clarity increases from 4, 8, and 16 channels to full spectrum. The matrix's pattern for cCI was closer to the pattern of cNH scores with eight-channel NBV than those with other conditions. The "scare" was the most difficult voice emotion to be recognized for the cCI across speakers and for the cNH listening to female speakers. The common error patterns were that cCI would misinterpret being scared as being happy (25.97%) when listening to female speakers and misinterpret being scared as being angry (37.42%) when listening to male speakers.

TABLE 3 List of sentences

Item#	English sentences (six syllables each)	Mandarin sentences
1	Her coat is on the chair.	她外套在椅子上。
2	The road goes up the hill.	這條路通山上。
3	They're going out tonight.	他們今晚要外出。
4	He wore his yellow shirt.	他穿了黃襯衫。
5	They took some food outside.	他們拿了一些食物去外面。
6	The truck drove up the road.	卡車開上路。
7	The tall man tied his shoes.	那男生綁緊鞋帶。
8	The mailman shut the gate.	郵差關上門。
9	The lady wore a coat.	那女孩穿著大衣。
10	The chicken laid some eggs.	雞生了幾顆蛋。
11	A fish swam in the pond.	魚在池裡游。
12	Snow falls in the winter.	冬天會下雪。

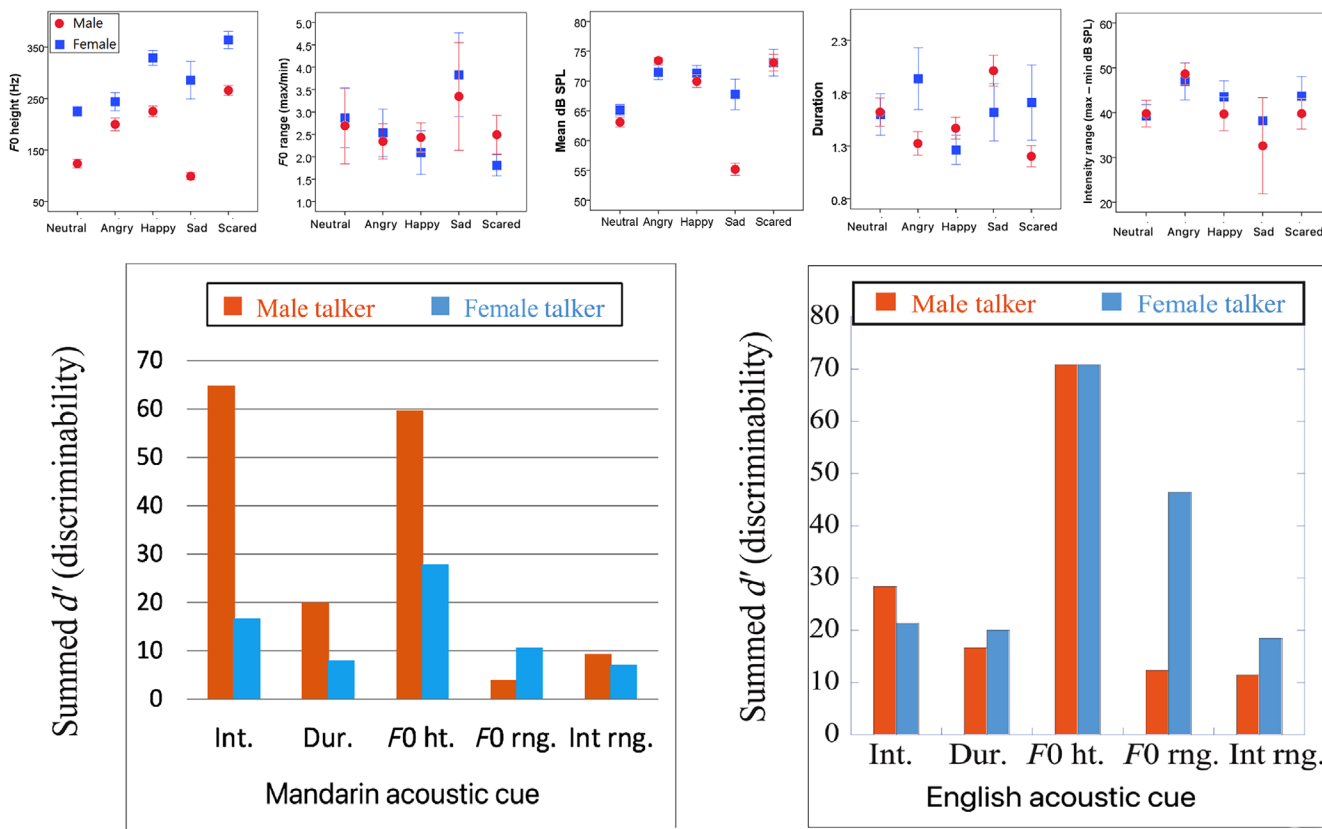


FIGURE 1 Results of acoustic analyses of male (red circles) and female (blue squares) speakers' utterances in five emotions (abscissa). For the top five panels, each panel corresponds to a different acoustic cue. Each point labeled in the y-axis represents the mean of all 12 sentences for each speaker, and error bars represent standard deviations. The bottom left panel is for the acoustic discriminability of Mandarin sentences, whereas the bottom right panel is for the acoustic discriminability of English sentences used in our previous study by Monita et al. SPL, Sound pressure level



FIGURE 2 Error pattern of voice emotion recognition. It shows the error patterns for the cCI and cNH groups of listeners, for the male and female speakers' sentences, and under each condition of spectral resolution tested. The cells are color-coded to represent the strength of the numerical values, but the actual values are also indicated. cCI, Cochlear-implanted children; cNH, normal-hearing children

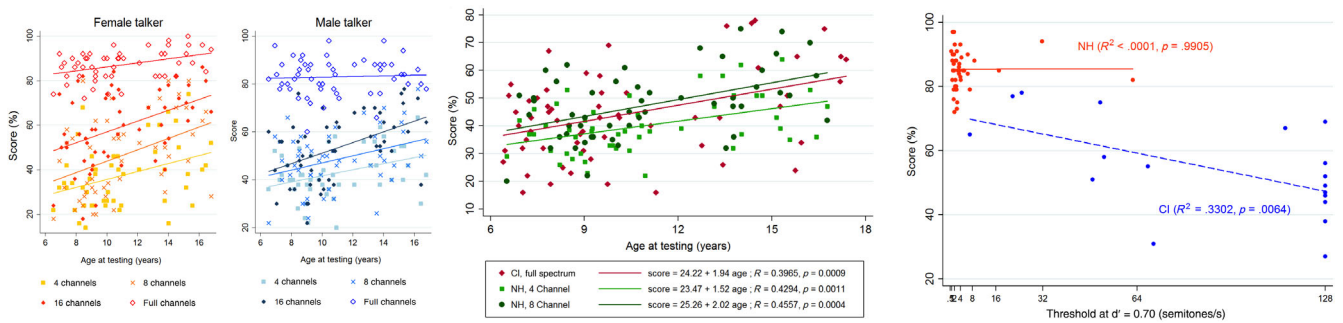


FIGURE 3 Mean voice emotion recognition score for cCI and cNH under different spectral degradation, speaker, and age. In the *left panel*, an LME analysis with RAU-transformed scores as the independent variable; age, condition (spectral resolution), and speaker as fixed effects; and subject-based random intercepts showed significant effects of age, $F(1, 51) = 21.42, p < .0001$; condition, $F(3, 51) = 2758.43, p < .0001$; speaker, $F(1, 51) = 6.26, p = .0156$; and a significant interaction between speaker and condition, $F(3, 51) = 8.49, p = .0001$. In the *central panel*, the average score of cCI with full-spectrum speech was close to the average score of cNH with eight-channel NBV speech. In the *right panel*, voice emotion recognition score as a function of F_0 threshold (semitone) revealed that the average performance across talkers for cCI could be predicted by their sensitivity to changes in F_0 (the thresholds extracted from the Weibull fits at a d' of 0.77; $R^2 = .3302; p = .0064$). cCI, Cochlear-implanted children; cNH, normal-hearing children; LME, linear mixed effects; NBV, narrowband vocoder; RAU, rationalized arcsine unit

3.2 | Group mean emotion recognition scores

3.2.2 | Full-spectrum speech

3.2.1 | Spectral degradation (for cNH)

A linear mixed-effects (LME) analysis with rationalized arcsine unit (RAU)-transformed scores as the independent variable; age, condition (spectral resolution), and speaker as fixed effects; and subject-based random intercepts showed significant effects of age ($p < .0001$), condition ($p < .0001$), speaker ($p = .0156$) and a significant interaction between speaker and condition ($p = .0001$; Figure 3). The cNH performance declined as the spectral resolution worsened.

An LME model with RAU-transformed scores as the independent variable; age at test, group (cNH or cCI), and speaker (male or female) as fixed effects; and subject-based random intercepts showed significant effects of age ($p = .0003$), group ($p < .0001$), and speaker ($p = .0003$), and a marginally significant interaction between age and group ($p = .0340$) on mean emotion recognition cores (Figure 3). The mean emotion recognition score of the cNH group was significantly better than the cCI. The female speakers' vocal emotions were more easy to be recognized; this difference was most apparent for the cCI group.

3.2.3 | Comparison between cCI and cNH

The cCI group showed a range of performance (including age dependency) like that of cNH attending to four-channel and eight-channel noise-vocoded speech (Figure 3). The average score of cCI with full-spectrum speech was close to the average score of cNH with eight-channel NBV speech.

3.3 | Sensitivity to *F0* changes

A large variability in the pitch sensitivity among implanted children was observed. Figure 3 (right panel) revealed that the average performance for voice emotion recognition across speakers for cCI could be predicted by their sensitivity to changes in *F0* ($r^2 = .3302$; $p = .0064$). However, this is not suggestive for cNH listening the sentences with full spectrum. Moreover, there was no significant effect for the age at implant ($p = .7552$), age at test ($p = .5998$), and duration of CI experience ($p = .7364$) on the task of discrimination of *F0* changes.

4 | DISCUSSION

Acoustic analysis of the Mandarin testing sentences revealed a substantial difference in the pattern of the summed discriminability indices for different cues compared with the English testing sentences used in our previous study¹⁶ (Figure 1). Happy was spoken with the greatest *F0* range and mean *F0* height in Chatterjee et al.'s study; however, scared was spoken with the greatest mean *F0* height and sad with the greatest *F0* range in the present study.

The discriminability measure (d') in the present study showed that *F0* height is the acoustic characteristics that contain the critical information, whereas *F0* range could additionally help for female voices and mean intensity for male voices. In the previous study by Chatterjee et al., the male speakers' sentences contained more information in the mean intensity patterns, whereas the female speakers' sentences contained more information in the *F0* range and the intensity range.¹⁶

The error patterns of voice emotion recognitions revealed large variability for cCI. A visual inspection of the patterns reveals that for cNH, the matrices become more and more diagonally dominant as spectral clarity increases. The diagonally dominant pattern observed for cCI is similar to that for four and eight-channel NBV speech observed in cNH. The error patterns of voice emotion recognitions for Mandarin-speaking cCI are not the same as English-speaking cCI. For example, the "scared" was the most difficult voice emotion to be recognized for the cCI across speakers for Mandarin-speaking cCI. Meanwhile, the most difficult voice emotion to be recognized for the English-speaking cCI was "scared" for male speakers and "neutral" for the female speakers.

In general, both cCI and cNH groups in full-spectrum and NBV condition obtained higher voice emotion recognition scores when listening to female speakers than when listening to male speakers. This difference was most apparent for the cCI group. This is inconsistent

with our previous study for the English-speaking peer.¹⁶ More information in the *F0* range is noted in the female speakers' sentences, whereas more information in the mean intensity patterns and duration is noted in the male speakers' sentences. This may suggest that *F0* is the primary cue used for voice emotion recognition. Nevertheless, children might recognize voice emotion based on secondary cues (such as intensity and duration) other than *F0* ranges for cCI or degraded NBV for cNH because *F0* cues are very severely degraded in CI and NBV with four or eight channels. Studies investigating music emotion processing found that CI users depend on tempo rather than pitch in the processing of musical emotion.²⁴⁻²⁷ The present study suggests a similar auditory processing strategy for emotion,²⁸ by increased reliance on cues such as intensity and duration that are closer to tempo-based aspects of music, for CI users compared to NH listeners.

Some cCI in this study could achieve high scores of emotion recognition. It would be interesting to investigate the underlying auditory emotion processing strategy for those cCI exhibiting high performance in this study. Although the *F0* cues are severely degraded in CI and NBV with four or eight channels, they might possess an unaccounted method to interpret *F0* information.²⁹

Participants' age at test has significant effect on voice emotion recognition as noted for both the cCI and cNH with degraded NBV in this study. However, the age of implantation did not show an effect on CI children's performances, suggesting that the effect is genuinely developmental in nature. We suppose that brain maturation plays a role in voice emotion recognition for cCI.

A tonal language benefit in pitch perception for children with CI has been reported in the literature.³⁰ Present results further revealed that a high sensitivity to changes in *F0* predicted a better performance of emotion recognition across Mandarin speakers cCI ($R^2 = .3302$; $p = .0064$). However, there was no significant effect for the age at implant, age at test, and duration of CI experience on the task of discrimination of *F0* changes. This suggests that, in addition to a psychological representation of *F0*, brain plasticity would also integrate other secondary auditory cues. Together with the positive effect of age at test for cCI on emotion recognition in present study, cCI might grow up with developed cognitive systems and adapted alternative ways to process auditory emotion.³¹ We suppose that improved sensitivity of tempo and intensity changes might be a main part of the development of cognitive systems for auditory emotion in cCI.

5 | CONCLUSION

As a result of device limitation in prosody processing, Mandarin-speaking cCI showed deficits in voice emotion recognition. Mandarin-speaking cCI performed comparably with cNH listening to spectral degraded speech, suggesting that cCI may have sufficiently developed adaptive strategies to interpret emotion from degraded auditory signals. Better pitch discrimination ability came with better voice emotion recognition. Besides the *F0* cues, cCI adapted their voice emotion recognition to rely more on secondary cues such as intensity and duration. Although cross-culture differences existed for the acoustic features of voice emotion, Mandarin-speaking cCI and their English-speaking cCI

peer exhibited a positive effect between age at test on emotion recognition, suggesting the learning effects or possibly maturation effects. Therefore, further device/processor development to improve the presentation of FO information and more rehabilitative efforts are needed to improve the transmission and perception of voice emotion.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ORCID

Yung-Song Lin  <https://orcid.org/0000-0002-2758-5163>

Charles J. Limb  <https://orcid.org/0000-0003-3109-1550>

REFERENCES

- Bryant G, Barrett HC. Vocal emotion recognition across disparate cultures. *J Cogn Cult*. 2008;8(1-2):135-148. doi:10.1163/156770908X289242
- Bryant GA. The evolution of human vocal emotion. *Emot Rev*. 2021;13(1):25-33. doi:10.1177/1754073920930791
- Cosme G, Tavares V, Nobre G, et al. Cultural differences in vocal emotion recognition: a behavioural and skin conductance study in Portugal and Guinea-Bissau. *Psychol Res*. 2021. doi:10.1007/s00426-021-01498-2 [published Online First].
- Planalp S. Varieties of cues to emotion in naturally occurring situations. *Cogn Emot*. 1996;10(2):137-154. doi:10.1080/026999396380303
- Xin L, Fu QJ, Galvin JJ 3rd. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif*. 2007;11(4):301-315. doi:10.1177/1084713807305301
- Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. *J Acoust Soc Am*. 2012;131(2):1307-1314. doi:10.1121/1.3672697
- Waaramaa T, Kukkonen T, Mykkanen S, Geneid A. Vocal emotion identification by children using Cochlear implants, relations to voice quality, and musical interests. *J Speech Lang Hear Res*. 2018;61(4):973-985. doi:10.1044/2017_jslhr-h-17-0054
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. *Nature*. 1991;352(6332):236-238.
- Wilson BS, Dorman MF. The surprising performance of present-day cochlear implants. *IEEE Trans Biomed Eng*. 2007;54(6):969-972.
- Peng SC, Tomblin JB, Cheung HT, Lin YS, Wang LS. Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear*. 2004;25(3):251-264. doi:10.1097/O1.Aud.0000130797.73809.40
- Deroche MLD, Kulkarni AM, Christensen JA, Limb CJ, Chatterjee M. Deficits in the sensitivity to pitch sweeps by school-aged children wearing Cochlear implants. *Front Neurosci*. 2016;10:73. doi:10.3389/fnins.2016.00073
- Peng SC, Lu HP, Lu N, Lin YS, Deroche MLD, Chatterjee M. Processing of acoustic cues in lexical-tone identification by pediatric Cochlear-implant recipients. *J Speech Lang Hear Res*. 2017;60(5):1223-1235. doi:10.1044/2016_jslhr-s-16-0048
- Lin YS, Lu HP, Hung SC, Chang CP. Lexical tone identification and consonant recognition in acoustic simulations of cochlear implants. *Acta Oto-Laryngologica*. 2009;129(6):630-637. doi:10.1080/00016480802032793
- Deroche MLD, Lu HP, Lin YS, Chatterjee M, Peng SC. Processing of acoustic information in lexical tone production and perception by pediatric Cochlear implant recipients. *Front Neurosci*. 2019;13:639. doi:10.3389/fnins.2019.00639
- Paulmann S, Uskul AK. Cross-cultural emotional prosody recognition: evidence from Chinese and British listeners. *Cogn Emot*. 2014;28(2):230-244. doi:10.1080/02699931.2013.812033
- Chatterjee M, Zion DJ, Deroche ML, et al. Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear Res*. 2015;322:151-162. doi:10.1016/j.heares.2014.10.003
- Barrett KC, Chatterjee M, Caldwell MT, et al. Perception of child-directed versus adult-directed emotional speech in pediatric Cochlear implant users. *Ear Hear*. 2020;41(5):1372-1382. doi:10.1097/aud.0000000000000862
- Hoff E. The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev*. 2003;74(5):1368-1378. doi:10.1111/1467-8624.00612
- Davis-Kean PE. The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *J Fam Psychol*. 2005;19(2):294-304. doi:10.1037/0893-3200.19.2.294
- Wechsler D. *Manual for the Wechsler Abbreviated Intelligence Scale (WASI)*. The Psychological Corporation; 1999.
- Lu L, Liu H. *The Peabody Picture Vocabulary Test-Revised in Chinese*. Psychological Publishing; 1998.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995;270(5234):303-304.
- Boersma P, Van Heuven V. Speak and unSpeak with PRAAT. *Glott International*. 2001;5(9/10):341-347.
- Hopyan-Misakyan TM, Gordon KA, Dennis M, Papsin BC. Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychol*. 2009;15(2):136-146.
- Volkova A, Trehub SE, Schellenberg EG, Papsin BC, Gordon KA. Children with bilateral cochlear implants identify emotion in speech and music. *Cochlear Implants Int*. 2013;14(2):80-91.
- Shirvani S, Jafari Z, Sheibanizadeh A, Zarandy MM, Jalaie S. Emotional perception of music in children with unilateral cochlear implants. *Iran J Otorhinolaryngol*. 2014;26(77):225-233.
- Giannantonio S, Polonenko MJ, Papsin BC, Paludetti G, Gordon KA. Experience changes how emotion in music is judged: evidence from children listening with bilateral cochlear implants, bimodal devices, and normal hearing. *PLoS One*. 2015;10(8):e0136685.
- Paquette S, Ahmed GD, Goffi-Gomez MV, Hoshino ACH, Peretz I, Lehmann A. Musical and vocal emotion perception for cochlear implants users. *Hear Res*. 2018;370:272-282. doi:10.1016/j.heares.2018.08.009
- Deroche MLD, Felezeu M, Paquette S, Zeitouni A, Lehmann A. Neurophysiological differences in emotional processing by Cochlear implant users, extending beyond the realm of speech. *Ear Hear*. 2019;40(5):1197-1209. doi:10.1097/aud.0000000000000701
- Deroche ML, Lu H-P, Kulkarni AM, et al. A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Sci Rep*. 2019;9(1):1-12.
- Gilbers S, Fuller C, Gilbers D, et al. Normal-hearing Listeners' and Cochlear implant Users' perception of pitch cues in emotional speech. *I-Perception*. 2015;6(5):19. doi:10.1177/0301006615599139

How to cite this article: Lin Y-S, Wu C-M, Limb CJ, et al. Voice emotion recognition by Mandarin-speaking pediatric cochlear implant users in Taiwan. *Laryngoscope Investigative Otolaryngology*. 2022;7(1):250-258. doi:10.1002/lio2.732