

# SCIENTIFIC REPORTS



OPEN

## A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children

Mickael L. D. Deroche<sup>1</sup>, Hui-Ping Lu<sup>2</sup>, Aditya M. Kulkarni<sup>3</sup>, Meredith Caldwell<sup>4</sup>, Karen C. Barrett<sup>4</sup>, Shu-Chen Peng<sup>5</sup>, Charles J. Limb<sup>4</sup>, Yung-Song Lin<sup>2</sup> & Monita Chatterjee<sup>3</sup>

In tonal languages, voice pitch inflections change the meaning of words, such that the brain processes pitch not merely as an acoustic characterization of sound but as semantic information. In normally-hearing (NH) adults, this linguistic pressure on pitch appears to sharpen its neural encoding and can lead to perceptual benefits, depending on the task relevance, potentially generalizing outside of the speech domain. In children, however, linguistic systems are still malleable, meaning that their encoding of voice pitch information might not receive as much neural specialization but might generalize more easily to ecologically irrelevant pitch contours. This would seem particularly true for early-deafened children wearing a cochlear implant (CI), who must exhibit great adaptability to unfamiliar sounds as their sense of pitch is severely degraded. Here, we provide the first demonstration of a tonal language benefit in dynamic pitch sensitivity among NH children (using both a sweep discrimination and labelling task) which extends partially to children with CI (i.e., in the labelling task only). Strong age effects suggest that sensitivity to pitch contours reaches adult-like levels early in tonal language speakers (possibly before 6 years of age) but continues to develop in non-tonal language speakers well into the teenage years. Overall, we conclude that language-dependent neuroplasticity can enhance behavioral sensitivity to dynamic pitch, even in extreme cases of auditory degradation, but it is most easily observable early in life.

**Tonal language benefit in pitch.** Speakers of a tonal language are continuously exposed to inflections in pitch, both rapid inflections within syllables to signify lexical tones<sup>1,2</sup> and slower inflections at the sentence level, to communicate prosody<sup>3</sup>. This places a strong informational emphasis on pitch which appears to influence its neural coding, both in the brainstem and cortex. For example, at the cortical level, pitch processing generally activates the right hemisphere, but could engage the left hemisphere when the pitch contours are linguistically relevant as in tonal languages<sup>4-6</sup>. Frequency following responses (FFRs) recorded at the brainstem can also reflect pitch processing<sup>7</sup>, and this is where researchers have sought evidence for a more robust encoding of periodicity in speakers of tonal language. A large body of work by Krishnan and his colleagues has repeatedly shown that while FFRs preserve pitch information of lexical tones in both tonal and non-tonal language speakers, they are more robust in the former (see review<sup>8</sup>).

The question arises as to how generalizable this enhancement in pitch coding is in speakers of tonal languages, e.g. how sensitive it is to the particular curvatures of pitch contours inherent to a given language. A first study<sup>9</sup> found no differences in pitch strength and pitch tracking accuracy derived from FFRs recorded in Mandarin and English speakers in response to linear fundamental frequency (F0) sweeps. Using iterated rippled noises, a second study<sup>10</sup> recorded FFRs in response to four stimuli: a prototypical and three artificial contours (an inverted version of tone 2, a linearly rising sweep, and a tri-linear approximation of tone 2). They found that Mandarin speakers exhibited higher pitch strength than English speakers only with the prototypical contour; population differences were lost for the three artificial contours. Also using iterated rippled noises, a third study<sup>11</sup> showed

<sup>1</sup>Centre for Research on Brain, Language and Music, McGill University, Rabinovitch House, 3640 rue de la Montagne, Montreal, Quebec, H3G 2A8, Canada. <sup>2</sup>Chimei Medical Center, Taipei Medical University, Tainan, Taiwan. <sup>3</sup>Auditory Prostheses and Perception Laboratory, Boys Town National Research Hospital, 555 N 30th Street, Omaha, NE, 68131, USA. <sup>4</sup>Department of Otolaryngology – Head and Neck Surgery, University of California San Francisco School of Medicine, San Francisco, CA, 94115, USA. <sup>5</sup>Center for Devices and Radiological Health, United States Food and Drug Administration, Silver Spring, Maryland, USA. Correspondence and requests for materials should be addressed to M.L.D.D. (email: [mickael.deroche@mcgill.ca](mailto:mickael.deroche@mcgill.ca))

that Mandarin speakers showed larger mismatch negativity responses (a cortical event-related response giving a look at sub-cortical, pre-attentive, processing) than English speakers to curvilinear pitch contours modelled after Chinese tones, but this population difference was lost when using linearly rising sweeps. Thus, these three studies arrived at the conclusion that the sharpening of pre-attentive processing of F0 is specific to contours that are ecologically relevant. A later study<sup>12</sup> tempered this view. On one hand, pitch tracking accuracy and pitch strength (similarly derived from brainstem FFRs in response to iterated rippled noises) were larger in Chinese than in English speakers, for sections of a lexical tone with rapid F0 changes, reinforcing the specificity of the tonal language advantage to particular curvatures in dynamic pitch contours. On the other hand, the effect was also observed with static F0 changes spanning a major third interval<sup>12</sup>. This musical interval was chosen to match the onset and offset of the lexical tone, and this is perhaps why the effect transferred to some degree outside of ecologically relevant stimuli (in Mandarin). Overall, it seems fair to conclude that, within neurophysiological studies looking at FFRs, the neural enhancement of pitch coding exhibited by speakers of tonal languages is most easily observable with ecologically relevant stimuli.

Another point of debate is the extent to which this neural enhancement translates into behavior. One of the earliest studies on this topic<sup>13</sup> already described the inconsistent status of the literature at the time on this question, and failed to find the hypothesized tonal language benefit in pure tone discrimination (in fact, observing the opposite, a tonal language deficit in pitch sensitivity). Another study<sup>14</sup> found no difference between Mandarin and English speakers in their ability to discriminate pure tones or pulse trains (while Mandarin speakers outperformed English speakers in Mandarin tone identification, as expected). In one of the FFR studies<sup>15</sup>, the neurophysiological enhancement did not translate into a perceptual advantage between Chinese and English speakers, and no correlation between F0 difference limens and FFR F0 magnitude was observed for these two groups. In contrast, two studies reported a tonal language benefit for pure tones in static discrimination or interval discrimination<sup>16,17</sup>. Using a very large sample through on-line testing, a tonal language benefit was revealed in detection of out-of-key incongruity<sup>18</sup>. Finally, a comprehensive study<sup>19</sup> reported an advantage for Cantonese over English speakers in a number of perceptual tasks (pitch discrimination, pitch speed, pitch memory, and melody discrimination). As pointed out by Bidelman and his colleagues<sup>19</sup>, differences in experimental tasks/designs are likely to account for some of these apparent discrepancies. First, the pitch of pure tones, pulse trains, or iterated rippled noises, may not be directly relevant to voice pitch or musical pitch. Second, some studies may not have tested sufficiently small frequency differences<sup>16</sup>, have included listeners from multiple language backgrounds<sup>16,17</sup> or have little audiological control<sup>18</sup>. Third, the cognitive abilities of participants could matter greatly in this phenomenon. For example Cantonese speakers exhibited superior working memory capacity (similar to English-speaking musicians) relative to non-musically trained English speakers<sup>19</sup>. This is a factor that could be particularly problematic with smaller-sample studies. One factor that has not been considered in previous studies is the developmental trajectory of brain plasticity. Given that younger brains are more plastic and yet also still developing cognitive skills, the tonal language advantage for pitch sensitivity may be more observable among children than among adults. Adults and older adolescents are likely to compensate for a lack of the tone-language advantage with more advanced cognitive skills and greater experience with diverse auditory inputs.

**Cochlear implant drawback for pitch.** Cochlear implants (CI) are devices implanted surgically that allow people with profound hearing loss to recover hearing to some degree. Despite stringent limitations in spectral resolution - among other aspects of signal degradation - CIs generally provide enough auditory information for speech to be intelligible, as long as the background environment is relatively quiet<sup>20</sup>. They achieve this feat by delivering electrical pulses to the cochlea that are modulated as a function of the acoustic input captured by an external microphone. Envelope-based coding strategies attempt to reproduce the modulations of temporal envelopes extracted from different frequency bands of the acoustic signal, thus recovering a sense of articulation that is critical to intelligibility<sup>21,22</sup>. Such strategies were designed for speech perception, but not for the complex harmonic pitch of the human voice or of musical instruments. This does not mean that pitch is impossible to perceive through CIs but rather, that it has to be retrieved by cues that are largely sub-optimal, resulting in a percept that is not as salient as in normal hearing<sup>23-25</sup>.

At first sight, it may seem counter-intuitive to study the possibility of a tonal language advantage among CI users given their limitations in pitch perception. But CI users have proven very useful to our understanding of normal hearing, much in the same way as investigating disease improves our understanding of health. More precisely, the common interpretation of the tonal language advantage is that it sharpens and strengthens the fine-grained representation of periodicity (as reflected by FFRs analysis). Most CI users (using envelope-based coding strategies) lack this fine-grained representation, and therefore there should not be any tonal language advantage within this population for pitch coding per se. If we were to find such an effect in CI users, this would imply an alternative mechanism for this phenomenon.

**Goal of the present study.** All studies aforementioned (section A) used adults only. To our knowledge, there is currently no data on the developmental trajectories of the hypothesized advantage (for pitch perception) of speaking a tonal language. As the phenomenon is directly related to language experience, it is a reflection of sensory neuroplasticity. As neuroplasticity changes dramatically in childhood, it is crucial to examine the effect across the developmental years. Thus, the main question addressed in the present study is the extent to which the tonal language effect holds within pediatric populations. In principle, younger brains should be more plastic. Thus, any neural enhancement in pitch coding should be more transferable to ecologically irrelevant pitch contours in developing children, while adults' brains might be more sharply tuned to the exact curvatures of pitch contours occurring within the tonal environment. From this standpoint, one may hypothesize that the tonal language benefit would be *more easily observable* in children than in adults. One reason this might not happen is if the developing brain of a child speaking a tonal language had not fully specialized yet to process lexical tones

	Chronological age mean (std.) [min – max]	Age at implantation mean (std.) [min – max]	Duration of CI experience mean (std.) [min – max]	Age at profound hearing loss mean (std.) [min – max]
NH – US (n = 44)	11.0 (2.8) [6.1–18.1]			
NH – Taiwan (n = 53)	11.0 (2.9) [6.9–16.8]			
CI – US (n = 52)	12.5 (3.3) [7.8–19.5]	2.8 (2.8) [0.3–14.0]	9.7 (3.7) [0.5–16.6]	0.7 (2.0) [0.0–12.0]
CI – Taiwan (n = 45)	10.5 (3.3) [6.6–17.2]	2.9 (1.8) [1.0–12.2]	7.6 (3.4) [1.2–15.2]	1.0 (0.7) [0.0–2.3]

**Table 1.** Demographics of the four groups of children.

as efficiently as adults, but this seems unlikely since typically developing children acquire mastery of lexical tones very early, before five years of age<sup>26,27</sup>. In the case for children with CIs, their brain is certainly not as specialized as their NH peers to process lexical tones<sup>28,29</sup>, but it has to adapt to a larger range of artificial/distorted sounds compared to NH children. So, in this sense, neuroplasticity may be even more important for children with CIs than for children with NH.

A second factor that may play a role is cognitive development. Adults may be able to compensate for their brain rigidity with their more advanced cognitive abilities. Thus, adult speakers of tonal and non-tonal languages may perform a pitch processing task using different mechanisms and skills than children of tonal and non-tonal languages. As in adults (previous studies in section A), we would expect to see minimal differences in our pitch sensitivity tasks between older adolescents who speak tonal or non-tonal languages, and we hypothesize that the difference would be greater for the younger children.

In a previous study<sup>30</sup> involving more than a hundred children, we investigated this very question and found no difference in F0 discrimination abilities between NH Taiwanese and NH Americans or between Taiwanese CI users and Americans CI users, but a large deficit was exhibited by the CI populations in both countries. This result provided no support for the tonal-language plasticity hypothesis in children. We reasoned that our static F0 discrimination task did not tap into the neural stages that were hypothetically sharpened by tonal language environments, because (1) broadband complex tones with static F0 such as those used in our studies are too remote from ecologically-relevant curvilinear tones, and (2) the effect could be highly task-dependent (present in an identification or labeling task; absent in a discrimination task, as noticed earlier<sup>14,31</sup>). Here, we revisited the hypothesis of a tonal language benefit in a behavioral study, using both a labelling and a discrimination task focused on dynamic F0 processing. In an earlier report that included only English-speaking listeners<sup>32</sup>, we used 300-ms long broadband harmonic complex sweeps with a range of linear slopes from very-shallow to very-steep. As expected, we found substantial deficits in F0-sweep sensitivity by children and adults wearing CIs compared to their NH peers. Interestingly, we also found age-related differences: adults outperforming children, and older children outperforming younger children. However, these differences were largely common to NH and CI subjects, suggesting that the role of cognitive development might not interact substantially with hearing status in this task. We now turn to the role of language-dependent plasticity, which might vary across age and hearing status.

We address the following critical questions: can a tonal language environment sharpen sensitivity to linearly rising/falling F0 in children with NH? If so, at what age? And does this benefit occur in their peers with CIs? Differences between the two tasks should reveal something about the nature of the hypothesized benefit. If speaking a tonal language acted as an enhancer of the internal representation of F0 (e.g. as reflected by the FFR F0 magnitude), it should provide benefits in both the labelling and the discrimination tasks. In contrast, if speaking a tonal language acted more at an abstract level (e.g. extraction of coarse features that are linguistically relevant), it could provide a specific advantage in the labelling task. This seems plausible since Mandarin-speaking listeners may naturally process pitch in a dynamic context - for instance, as rising versus falling - and hence may find the present tasks more intuitive than English-speaking listeners who generally define pitch on a scale going from low to high in a musical context. English speakers would encounter dynamic pitch in speech in the context of prosodic cues, but those are generally slower, occurring over the course of a sentence, and perhaps less crucial to comprehending the meaning of utterances than lexical tones, which are an integral component of words in Mandarin.

## General Methods

**Subjects.** Four groups of children participated. They included 44 Americans with NH (21 of whom had previously been reported<sup>32</sup>), 53 Taiwanese with NH, 52 Americans with CI (23 of whom had previously been reported<sup>32</sup>), and 45 Taiwanese with CI. The chronological age of all participants varied from 6.1 to 19.5 years. Details for each group are provided in Table 1. A large majority of the children with CI (45 Americans and all Taiwanese) were profoundly deaf at birth or within their first year of life. Age at implantation varied from 4 months to 14 years. Their duration of CI experience varied from 6 months at minimum up to 16.6 years. A minority of Americans (10) were unilaterally implanted (3 on the left side, 7 on the right), while 45 were implanted on both sides. In contrast, a majority of Taiwanese (39) were unilaterally implanted (16 on the left side, 23 on the right), while only 6 of them were implanted on both sides. Among the children implanted unilaterally, about half of them had sufficient residual hearing to wear a hearing aid on the contralateral ear. All children with CI, however, were tested on one ear only. Children with two implants were asked to unplug the most recent implant. For children with a single implant, ear-foam plugged the contralateral ear and any hearing aid was removed.

A large majority of Taiwanese (39) had a *Cochlear* device. Only two children had an *Advanced Bionics* device, and four had a *Med-El* device. American children were roughly split between *Cochlear* (24) and *Advanced Bionics* (26) devices, with only two children wearing a *Med-El*. All stimulation strategies were envelope-based (mostly ACE strategy for Nucleus 24, N5 and N6; mostly HiRes strategy for the Clarion, Naida, or Neptune; and the standard strategy for the Sonata and Opus2). Implanted children did not switch to a music program in their processor for this study: they used their CI as it was clinically assigned to them on a daily basis.

**Stimuli.** The stimuli were identical to those presented in the earlier report<sup>32</sup>. Harmonic complexes were generated with all their partials up to the Nyquist frequency, with equal amplitude and in sine phase. The flat spectral envelope of these stimuli leaves little room for NH listeners to focus on individual harmonics within a particular frequency region. Instead, listeners must derive a dominant pitch percept by integrating periodicity information across spectral channels. For CI users, the same cannot be said with any degree of certainty: some subjects could be sensitive to a particular region of the cochlea with a better electrode-neuron interface or with higher number of healthy neurons to relay the auditory information higher up. Different processing strategies (variable across devices and manufacturers) could also promote listening to certain channels. Thus, it remains unclear how individual CI users derive their percept of global pitch from these stimuli (whether it is from envelope periodicities, or spectral centroid, or an even rougher comparison of place pitch across adjacent electrodes). However, for both NH and CI listeners, the percept of spectral edge pitch was eliminated by low-pass filtering at 10 kHz using Butterworth sixth-order filter with a slope of  $-30$  dB per octave. The duration of the complexes was fixed at 300 ms to retain relevance to syllabic durations in speech, gated with ramps of 30 ms. The interval between stimuli was also set at 300 ms. In order to obtain a view of the entire psychometric function for every subject, sweeps had to range from very shallow (for NH subjects) to very steep (for CI subjects). Thus, the F0 of the complex could vary in a logarithmic space with rates of  $\pm 0.5$ , 1, 2, 4, 8, 16, 32, 64, and 128 semitones per second, and the scale was adjusted for individual subjects for performance to vary from chance to ceiling. Sweeps with opposite direction always shared the same F0 range, such that judgements had to rely on the evolution of the pitch percept over time rather than the overall pitch range that was covered. This was critical because CI users were expected to perceive many of the sweeps as flat, and a natural consequence is to start listening to the height of the pitch elicited rather than where it is heading towards. In the same logic, the base F0 was roved across trials between 100 and 150 Hz, to discourage children from relying on F0 range. All stimuli were generated at 65 dB SPL and presented with  $\pm 3$  dB level roving.

**Protocol.** After explaining the protocol and obtaining informed written consent from children and parents, the participants were invited to sit in the auditory booth and practice blocks commenced. The first task was a forced-choice procedure with a single interval and two alternatives (1I-2AFC). Children listened to a single F0-sweep and reported whether the pitch was rising or falling. The second task was a forced-choice procedure with three intervals and two alternatives (3I-2AFC). One F0-sweep was played as a reference, followed by two others, either identical or in opposite direction (randomly placed) at the same rate. Children reported the interval that sounded *different* from the reference. Before testing took place, practice blocks were presented without level roving, and with 20 trials to evaluate performance with the most extreme sweeps, namely 128 semitones/sec. The test was only initiated provided that children obtained at least 80% correct, averaged over the two directions (up or down). A test block usually contained 140 trials (7 rates by 2 directions, tested 10 times each, all shuffled). The seven rates ranged from 0.5 to 32 semitones/sec for NH listeners, and ranged from 2 to 128 semitones/sec for CI listeners. Once this first block was completed, the experimenter looked at the data collected and adjusted the scale of sweep rates for subsequent blocks if performance was too close to floor or ceiling, such that enough data could be obtained in the middle of the psychometric function.

The interface consisted of animated cartoons synchronized with presentation of the sounds and buttons on the screen<sup>33</sup>. In each trial, reaction time (RT) was recorded, although subjects were not aware of it. Feedback was provided via smileys and the experimenter tried to keep each child motivated by challenging him/her to win points and bonuses (not used for analysis). Experimental sessions lasted about one hour with short breaks between blocks. All children were paid for their participation. This study was approved by the Institutional Review Board of Johns Hopkins, BTNRH, UCSF, Chi Mei Medical Center and the Chang Gung Memorial Hospital, in accordance with principles expressed in the Declaration of Helsinki.

**Equipment and testing sites.** The study took place at five research facilities. Taiwanese data were collected at the Chi Mei Medical Center in Tainan (52%) and at the Chang Gung Memorial Hospital in Taoyuan (48%). American data were collected at Boys Town National Research Hospital in Omaha (48%), and at the School of Medicine of UCSF in San Francisco (30%), and at Johns Hopkins Hospital in Baltimore (22%). The effect of experimental site in the US as well as in Taiwan was tested in each task but never reach significance ( $p > 0.153$ ). The setups were broadly similar across sites. Stimuli were sampled at 44.1 kHz with a resolution of 16 bits. They were presented via a loudspeaker placed half a meter from the subject at 65 dB SPL. The loudspeaker (SB-1 Audio Pro at all sites in Taiwan, Grason Stadler GSI at BTNRH, and Sony SS-MB150H at UCSF and Johns Hopkins) was located in front of the subject, and the experimental interface was shown on a monitor, inside a sound-proof audiometric booth. Listeners gave their responses by touching a screen or using a mouse.

### Data Analysis

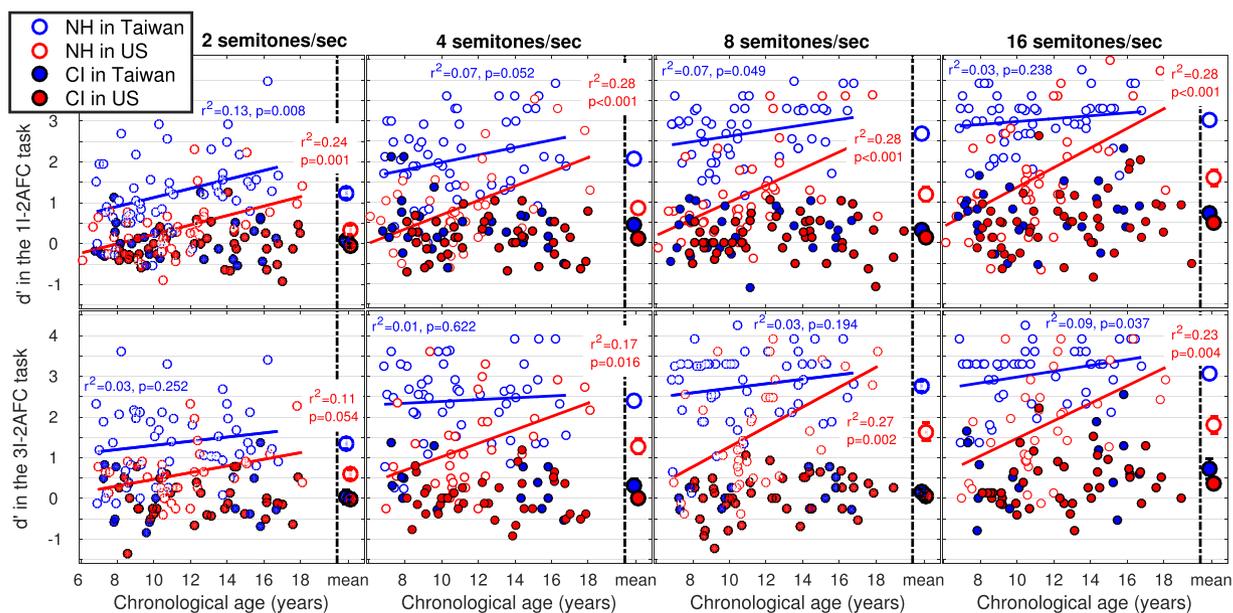
The first analysis compared performance across populations for a given type of pitch inflections. To this aim, performance was averaged over test blocks at a given rate, then translated into a number of hits and a number of false alarms, which could provide an estimate of  $d'$  and beta (i.e. bias in responding)<sup>34</sup>. When performance reached 100% over  $n$  trials, it was instead adjusted to  $100 \times (1 - 1/(2n))$ , in order to keep  $d'$  within finite values<sup>35</sup>.

Sweep rate (semitones/sec)	0.5	1	2	4	8	16	32	64
language			F(1,156) = 21.1 *p < 0.001	F(1,160) = 35.3 *p < 0.001	F(1,166) = 49.7 *p < 0.001	F(1,168) = 37.3 *p < 0.001		
hearing			F(1,156) = 49.4 *p < 0.001	F(1,160) = 81.6 *p < 0.001	F(1,166) = 209.3 *p < 0.001	F(1,168) = 157.7 *p < 0.001		
language × hearing			F(1,156) = 12.9 *p < 0.001	F(1,160) = 10.9 *p = 0.001	F(1,166) = 30.6 *p < 0.001	F(1,168) = 19.4 *p < 0.001		
NH-US vs. NH-Taiwan	t(60) = 0.9, p = 0.346	t(93) = 2.0, *p = 0.045	F(1,156) = 41.7 *p < 0.001	F(1,160) = 51.7 *p < 0.001	F(1,166) = 93.3 *p < 0.001	F(1,168) = 64.6 *p < 0.001		
CI-US vs. CI-Taiwan			F(1,156) = 0.4 p = 0.519	F(1,160) = 3.0 p = 0.087	F(1,166) = 1.0 p = 0.321	F(1,168) = 1.3 p = 0.260	t(75) = 4.6, *p < 0.001	t(75) = 3.5, *p = 0.001

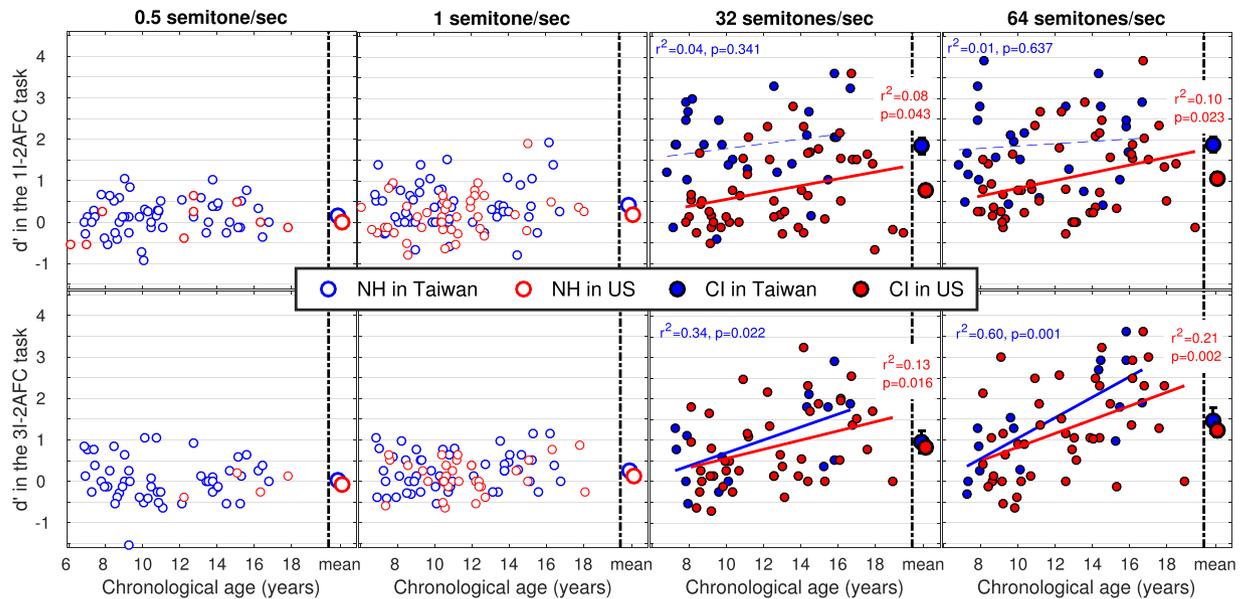
**Table 2.** Results of the statistical analysis of the  $d'$  data obtained in the labelling task, for isolated sweep rates (shown in the top panels of Figs 1 and 2).

Sweep rate (semitones/sec)	0.5	1	2	4	8	16	32	64
language			F(1,130) = 8.7 *p = 0.004	F(1,132) = 20.8 *p < 0.001	F(1,135) = 14.1 *p < 0.001	F(1,137) = 25.7 *p < 0.001		
hearing			F(1,130) = 49.7 *p < 0.001	F(1,132) = 111.0 *p < 0.001	F(1,135) = 163.7 *p < 0.001	F(1,137) = 138.9 *p < 0.001		
language × hearing			F(1,130) = 6.5 *p = 0.012	F(1,132) = 6.6 *p = 0.011	F(1,135) = 9.7 *p = 0.002	F(1,137) = 7.8 *p = 0.006		
NH-US vs. NH-Taiwan	t(52) = 0.5, p = 0.654	t(81) = 1.3, p = 0.194	F(1,130) = 22.2 *p < 0.001	F(1,132) = 36.8 *p < 0.001	F(1,135) = 33.7 *p < 0.001	F(1,137) = 43.8 *p < 0.001		
CI-US vs. CI-Taiwan			F(1,130) < 0.1 p = 0.808	F(1,132) = 1.5 p = 0.223	F(1,135) = 0.2 p = 0.696	F(1,137) = 2.0 p = 0.161	t(57) = 0.4, p = 0.660	t(57) = 0.7, p = 0.507

**Table 3.** Results of the statistical analysis of the  $d'$  data obtained in the discrimination task, for isolated sweep rates (shown in the bottom panels of Figs 1 and 2).



**Figure 1.**  $d'$  data across children for 300-ms sweeps of  $\pm 2$  (most-left)  $\pm 4$  (middle-left),  $\pm 8$  (middle-right), and  $\pm 16$  (most-right) semitones/sec, in two tasks where the child was asked to label the direction of a single sweep (top panels) or discriminate between sweeps of opposite direction (bottom panels). Means are on the right-hand side of each panel, and error bars represent one standard error. A higher  $d'$  reflects a more acute sensitivity. At these rates, none of the regressions reached significance among children with CIs, and age effects were stronger for NH children in the US than in Taiwan.



**Figure 2.**  $d'$  data across children for 300-ms sweeps of  $\pm 0.5$  (most-left)  $\pm 1$  (middle-left),  $\pm 32$  (middle-right), and  $\pm 64$  (most-right) semitones/sec, in the labelling (top) or discrimination (bottom) task. Means are on the right-hand side of each panel, and error bars represent one standard error. At the shallow rates, none of the correlations reached significance among NH children, and at steep rates, age effects were more consistent for implanted children in the US than in Taiwan.

An analysis of variance (ANOVA) with two between-subjects factors - *hearing status* and *language background* - was performed on the  $d'$  data, separately for each rate that was sufficiently common to NH and CI populations, namely 2, 4, 8, and 16 semitones/sec. Post-hoc comparisons were performed between children in the US and in Taiwan to evaluate the tonal language benefit in each hearing status (Tables 2 and 3). Too few CI subjects could operate at 0.5 and 1 semitone/sec, so differences between NH Taiwanese and NH Americans were assessed by an independent-samples t-test. Similarly, too few NH subjects had been tested on 32 and 64 semitones/sec (as it was too easy for them), and differences between Taiwanese and Americans among users of CI were assessed by an independent-samples t-test.

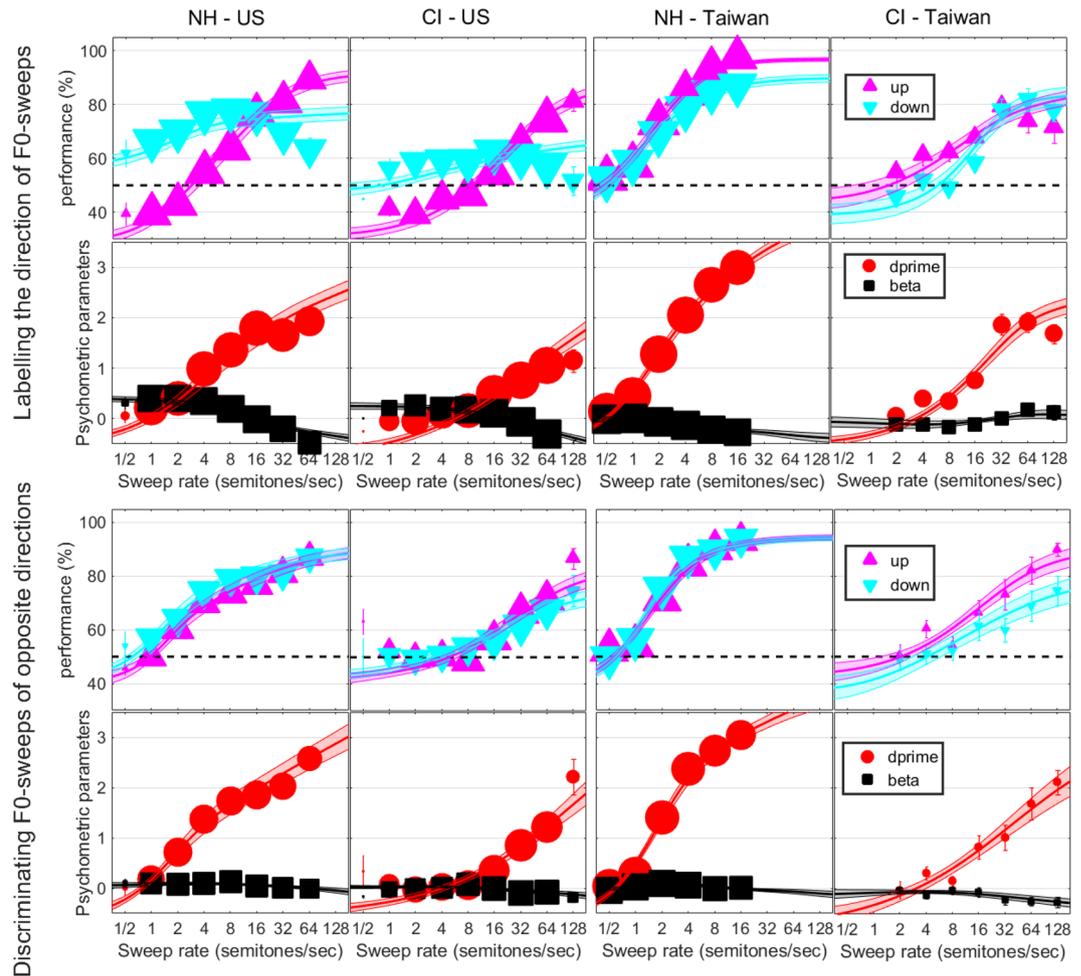
The second analysis was focused on comparing the rate of pitch inflections that was required for the different populations to reach a fixed level of discriminability, namely  $d' = 0.77$ . This normative value was chosen to ease comparisons with studies using adaptive staircases. To this aim, performance data were fitted (separately for up-sweeps and down-sweeps) with a Weibull function, using the maximum-likelihood technique<sup>36,37</sup>. The scale of sweep rates varied logarithmically (base 2). The fitting procedure was facilitated by Gaussian priors. The lower and upper bounds were probed around performance at, respectively, the shallowest and steepest rate available, and a standard deviation of 30%. There was no prior for the position of the inflection point but its slope was searched around a mean of 1 and a standard deviation of 3. Subsequently,  $d'$  fits were generated from the fits of up- and down-sweeps, enabling extraction of a threshold (expressed in semitones/sec) at exactly  $d'$  of 0.77 for each child. In this process, the performance data could reach 100% but the Weibull fit was limited at  $100 \times (1 - 1/(2n))$  to keep the resulting  $d'$  fit within finite values<sup>35</sup>. In a number of cases (18% of Americans with CI, 18% of NH Americans, 11% of Taiwanese with CI, and 1% NH Taiwanese), the maximum likelihood technique could not find any acceptable Weibull fit to the performance of down-sweeps because it exhibited a flat or non-monotonic pattern, an issue made transparent in our earlier study<sup>32</sup>. Occasionally, this also occurred for up-sweeps (8% of Americans with CI, 6% of Taiwanese with CI, 2% of NH Americans). Whenever this happened, the data were fitted with a straight line corresponding to the average performance across all rates tested. In those cases, the  $d'$  fit was thus primarily based on the Weibull fit obtained from the other sweep direction. Finally, thresholds were assigned to “chance” (and excluded from further analysis) whenever the  $d'$  fit could not exceed 0.77 by a sweep rate of 256 semitones/sec. To evaluate the measurable thresholds statistically, an ANOVA with two between-subjects factors - *hearing status* and *language background* - was performed, and post-hoc comparisons between children in the US and Taiwan tested the tonal language benefit in each hearing status (Table 4).

## Results

**Analysis of  $d'$ .** Figure 1 shows the  $d'$  data measured for each of the four populations in the labelling task (top panels) and the discrimination task (bottom panels). In both tasks, the results of the ANOVAs reveal significant main effects and interaction (Tables 2 and 3). The large deficits exhibited by children with CI were expected, and therefore interactions were explored for the effect of language background. Post-hoc tests revealed that NH Taiwanese outperformed NH Americans at 2, 4, 8, and 16 semitones/sec, and consistently in both tasks. This represents strong evidence for a tonal language benefit among NH children, with an effect size from 1.0 to 1.5 gain in  $d'$ . In contrast, there were too few differences between Taiwanese and Americans wearing CI at those rates.

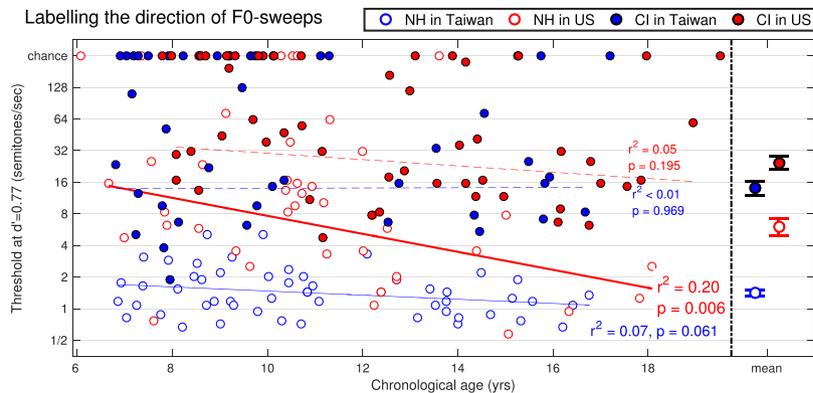
Task	1I-2AFC	3I-2AFC
language	$F(1,148) = 41.6$ * $p < 0.001$	$F(1,118) = 5.9$ * $p = 0.017$
hearing	$F(1,148) = 143.0$ * $p < 0.001$	$F(1,118) = 151.3$ * $p < 0.001$
language $\times$ hearing	$F(1,148) = 8.3$ * $p = 0.005$	$F(1,118) = 11.5$ * $p = 0.001$
NH-US vs. NH-Taiwan	$F(1,148) = 53.2$ * $p < 0.001$	$F(1,118) = 26.6$ * $p < 0.001$
CI-US vs. CI-Taiwan	$F(1,148) = 5.4$ * $p = 0.021$	$F(1,118) = 0.3$ $p = 0.558$

**Table 4.** Results of the statistical analysis of the thresholds extracted at  $d' = 0.77$  (shown in Figs 4 and 5).

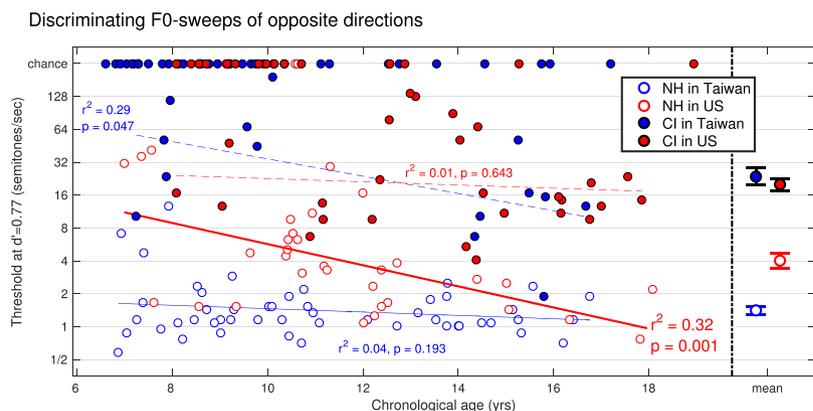


**Figure 3.** Psychometric functions expressed as percent correct for up- and down-sweeps and subsequently converted into  $d'$  values, in the 1I-2AFC labelling task (top) and the 3I-2AFC discrimination task (bottom), as a function of the rate of F0-sweeps expressed in semitones per second. Symbols represent the weighted mean and error bars indicate one weighted standard error of the mean. Note that the size of symbols relates to the weights on each condition reflecting the number of trials collected across all subjects. Lines and surfaces are the Weibull fits with one standard error.

Figure 2 shows the  $d'$  data measured for the two NH populations at very shallow rates (left panels) and for the two CI populations at very steep rates (right panels). As illustrated on the bottom panels, there was no tonal language benefit among NH children or among children with CI in the discrimination task (stats reported in Table 3). However, in the labelling task (top panels, with stats reported in Table 2), there was a tonal language benefit among children with CI at both 32 and 64 semitones/sec, and a small benefit among NH children at 1 semitone/sec (although it would not survive Bonferroni correction).



**Figure 4.** Individual thresholds across children and means on the right-hand side with one standard error, for the labelling task. Here, a better sensitivity is reflected by a lower threshold. Children who could not perform at this level of  $d'$  are reported as chance on the top-end of each panel.



**Figure 5.** Same as Fig. 4 for the discrimination task.

Linear regressions investigating the effects of chronological age revealed a striking asymmetry. For NH children in Taiwan, regression coefficients were not all significant between 2 and 16 semitones/sec (Fig. 1) and never explained more than 13% of the variance. In contrast, for NH children in the US, the effect of age was stronger: regression coefficients were consistent between 2 and 16 semitones/sec and explained up to 28% of the variance. This asymmetry was paralleled in the CI groups: at 32 and 64 semitones/sec, regression coefficients were significant for children in the US in both tasks but only in the discrimination task for children in Taiwan.

**Analysis of thresholds.** Figure 3 shows performance (in % correct) for up- and down-sweeps in each population and each task, along with the corresponding psychometric parameters,  $d'$  and  $\beta$ . This illustrates more clearly how  $d'$  improved with increasing sweep rates. In the discrimination task, performance started around 50% for shallow rates and increased monotonically as sweeps became steeper, with little bias with regard to the direction. This was not the case in the labelling task, in which performance differed from chance level at shallow rates. In the US more particularly, children were biased to respond “down” more often at shallow rates, and “up” more often at steep rates. This anomaly has been extensively discussed<sup>32</sup> and reflects that some listeners (here children, but this also applies to adults) judge a given sweep by comparison with many sweeps presented earlier. When sweeps hit distinct F0 ranges depending on their rate (which has to occur for very steep sweeps), the percept of F0 height becomes difficult to ignore, and in this respect Taiwanese children (both NH and CI) seem more immune than American children (see general discussion). Note that this bias can cause non-monotonicity of the psychometric function, which the Weibull fit cannot follow. Nonetheless, when combining both directions, the  $d'$  fits are in relatively good correspondence with the  $d'$  data, such that the threshold derived for a given child was already a good representation of his/her sensitivity.

Figures 4 and 5 show the thresholds, i.e. the values of sweep rate that would be required for each child to reach a common level of discriminability namely  $d'$  of 0.77. Note that there were some children who did not obtain 80% performance during practice and there were children who passed the criterion but provided data which were very close to chance such that a threshold was not measurable at  $d' = 0.77$ . All these subjects are shown here as “chance” on the top-end of each panel. It is clear that many children with CIs (of both language backgrounds) could not do these tests. Even with extremely steep sweeps of 128 semitones/sec (corresponding to a change of 3.2 octaves within 300 ms), many of them, especially the younger children, could not tell whether the pitch was rising

or falling, and discriminate between the two. This speaks to the considerable limitations in complex pitch coding currently offered by envelope-based devices. Among all participants who provided a measurable threshold, the statistical results (reported in Table 4) revealed consistent findings in both tasks: a main effect of language background, a main effect of hearing status, and an interaction. Delving further into the interaction, NH Taiwanese outperformed NH Americans in both tasks, and Taiwanese users of CI outperformed American users of CI in the labelling task only. In other words, this is qualitatively the same pattern that emerged from the first analysis.

The effect of chronological age was again evaluated with linear regressions. This factor accounted for 20% and 32% of the variance among NH Americans while it did not reach significance for NH Taiwanese. Among children with CIs, many could not perform the tasks, so age effects were inconsistent. Overall, this is also similar to the pattern that emerged from the first analysis.

Given the different slopes of age effects between the two NH populations, an additional analysis was performed to evaluate the effect of age on the tonal language *benefit*. To this aim, the age axis was split into 22 bins with a width of 2 years, equally spaced every 0.5 year. The thresholds for the children falling into a given bin were pooled together, and a mean and standard deviation was calculated. By computing the difference between the means of the two NH populations divided by their pooled standard deviation, an estimate of effect size was derived in each bin. A linear trend could well account for the progressive reduction in the tonal language benefit with chronological age ( $r^2 = 0.51$ ,  $p < 0.001$  in the labelling task;  $r^2 = 0.23$ ,  $p = 0.024$  in the discrimination task) which would have completely disappeared by age 20 and 22, respectively in each task.

**RT data.** For each trial, RT was stored separately for correct and incorrect responses. In each population and each task, averaged RT for correct responses decreased as sweep rate increased. The effect was best described by linear regressions, decreasing by several hundred milliseconds (decreasing by 100 ms at minimum and 600 ms at most) from the shallowest to the steepest rates. This pattern was largely expected and can be taken as evidence for diligent behavior<sup>38</sup>: children presumably tried to perform as well as possible because they took a bit more time to provide their responses when facing subtle cues.

**Probing other factors.** Thresholds obtained by children with CI were also examined as a function of age at implantation and duration of CI experience, but none of the correlations were significant, even when pooling the two populations together ( $p > 0.204$ ). Among other factors of interest, there were no differences between unilaterally and bilaterally implanted children in either task [ $F(1,59) = 2.3$ ,  $p = 0.133$ ;  $F(1,41) = 3.0$ ,  $p = 0.089$ ], no effect of the side of ear tested [ $F(1,59) = 0.1$ ,  $p = 0.722$ ;  $F(1,41) = 2.5$ ,  $p = 0.120$ ], and no interaction between the two [ $F(1,59) < 0.1$ ,  $p = 0.858$ ;  $F(1,41) = 3.6$ ,  $p = 0.064$ ]. The effect of contralateral hearing aid use did not reach significance [ $F(1,61) = 3.3$ ,  $p = 0.072$ ;  $F(1,43) = 0.4$ ,  $p = 0.538$ ], and there was no difference between devices built by different manufacturers [ $F(2,60) = 0.3$ ,  $p = 0.768$ ;  $F(2,42) < 0.1$ ,  $p = 0.962$ ].

## General Discussion

A clear tonal language benefit was observed among NH children. This was true for the ability to label the direction of a single sweep, as well as the ability to discriminate between sweeps of opposite direction. This was demonstrated first by the finding that NH Taiwanese obtained higher  $d'$  than their American peers across a very large diversity of sweeps, from F0s that rose/fell by only 60 cents to as much as 4.8 semitones within 300 ms (Fig. 1). Second, this was demonstrated by the finding that NH Taiwanese needed sweep rates of only 1.4 semitone/sec whereas their American peers needed sweep rates of 5.4 semitone/sec (on average across tasks) to achieve the same value of  $d'$  (Figs 4 and 5). Perhaps even more interesting is the fact that this tonal language benefit would not survive in adulthood. Indeed, chronological age had a stronger role for Americans than for Taiwanese. Presumably, Mandarin-speaking children learn early on in life that pitch inflections are critical to pay attention to, and this gives them an edge over young English-speaking children in dynamic-pitch tasks. Therefore, researchers should probe the early years of language development when seeking evidence for tonal language benefits.

What is also striking is that this tonal language benefit was absent in our previous study<sup>30</sup> measuring static F0 sensitivity, using otherwise similar design, equipment, and analysis. This means that the tonal language benefit may not generalize to pitch perception overall, at least in children. Rather, it seems to concern the sensitivity to continuous changes in pitch over time (as in speech) and not discrete ones (as in a melodic sequence of piano notes), providing no support for cross-domain generalization (speech to music). This finding also supports the general conclusion based on the literature that if tonal language benefits are to be found in pitch perception tasks, they are more likely present in tasks that are closely related to the demands of the tonal language.

The tonal language benefit is often interpreted in terms of enhancement in the internal representation of F0, largely owing to the number of studies that showed a stronger encoding of periodicity in brainstem FFRs of tonal language speakers (see section A of the introduction). The NH data in this study are in line with this standard interpretation. The lack of tonal language benefit among CI users in the discrimination task is also consistent with this framework, since F0 representation is severely degraded within this population. As such, CI users (regardless of their language background) must derive a pitch percept from a cue that is poorly coded in terms of spectral harmonic structure and in terms of temporal waveform<sup>24</sup>. There is much interest in understanding what exactly that cue is, whether it is primarily derived from temporal envelope periodicity<sup>23,39–43</sup>, or a rough spectral centroid constructed from place cues<sup>44–46</sup>, and whether different pitch percepts can be interchangeable<sup>47,48</sup>. One result, however, that achieves consensus in the field is that the brains of CI users can adapt to the signal degradations in the input to some extent. There is a plethora of clinical evidence for the benefits of early implantation<sup>49–58</sup>. Unfortunately, much of this clinical evidence is focused on speech and oral communication, and it is doubtful that neural plasticity in the early stages of development can overcome the limitations of CIs in terms of pitch. Both the present and previous results<sup>30</sup> cast a pessimistic light in this regard: neither age at implantation nor years of CI use had any impact on the sensitivity to static or dynamic F0, which remained overall poor. Here, thresholds

were about 16–32 semitones/sec (on average) and this effect size is actually underestimated because it is based on the children who could provide a measurable threshold at a  $d'$  of 0.77. Even with sweeps of 128 semitones/sec, 34% of children with CIs could not tell whether the pitch was rising or falling, and 44% could not discriminate between the two. Yet, most of these children were implanted before 4–5 years of age and had at least six years of experience with the device. Thus, we believe that it is rather unlikely that neuroplasticity will suffice to recover a sharp internal representation of F0 in children wearing cochlear implants. Note that this is not the first study to reach this conclusion: although there is a quick adaptation following device activation, children seem to suffer largely from the same difficulties as adults (wearing cochlear implants) in pitch-related tasks such as emotion recognition, prosody perception/production, and lexical tone identification<sup>28,59–67</sup>.

From this standpoint, the discovery that Taiwanese children with CIs outperformed their US counterparts only in the labelling task is of great interest. This implies that there is an additional explanation for this phenomenon, potentially unrelated to the internal representation of F0 and specific to what happens in the brain during “labelling”. One particularity of the labelling task (discussed in depth<sup>32</sup>) is that a single sound is presented without explicit reference to a common pitch range. When a sweep sounds roughly flat, the listener may respond based on the pitch range, i.e. “going up” for a high-pitch range and “going down” for a low-pitch range. That is, listeners may confuse the meaning of pitch direction with that of pitch height. This is not a strategy that is only recruited by CI users; NH listeners fall back on it too, provided that the sweeps are sufficiently shallow to be perceived as flat. However, the problem is somewhat exacerbated for CI users because they cannot perform the task until the sweeps are very steep. Sweep rates of 64 or 128 semitones/sec necessarily elicited a high pitch range (either from the beginning or the end of the sound). Since the sweeps were only 300-ms long, children with CI could perceive them as one diffuse high range without any motion to them, which contrasted with the relatively low range used for other (shallower) sweeps. In other words, in the 1I-2AFC, listeners accumulate precedent trials as internal references which bias the decision made on a given trial. This problem does not arise in the 3I-2AFC task because this reference is made explicit by having the three sweeps covering the same pitch range (no bias between up/down in the bottom panels of Fig. 3). Therefore, we speculate that the advantage exhibited by Taiwanese over American children wearing CIs in the labelling task is in fact a conceptual advantage, derived from a greater familiarity with the task demands. Through their natural and intensive exposure to Mandarin, the Taiwanese children may have disambiguated the notions of pitch height and pitch direction better than American children, rendering them more immune to the bias of the height of the pitch range covering steep sweeps. In fact, from multidimensional scaling studies across languages, there is already some evidence that speakers of tonal languages weigh the direction dimension more heavily than English speakers, whereas English speakers weigh the height dimension more heavily than tonal language speakers<sup>68</sup>, and this has repercussions in the strategies that people use when learning a new language. For example, English speakers learnt to categorize Cantonese tones by relying heavily on their height whereas Mandarin speakers learnt to categorize them by relying more on their direction<sup>69</sup>. Even though this study examined young adults, none of them had any familiarity with Cantonese, so it bears some similarity to children acquiring their native language. This supports a different interpretation of the tonal language benefit, namely that children pay particular attention to the pitch dimension that is most phonetically relevant in their language, which in this study favored the Taiwanese whereas it favored the Americans in our previous study<sup>30</sup>. This difference alone in dimension-weighting may be sufficient in accounting for the diametrically opposite pattern of results in our two studies.

## Summary

We measured the full psychometric functions in four groups of children (Taiwanese and Americans, with normal hearing or wearing cochlear implants) for the sensitivity to linear glides in complex pitch (broadband harmonic complexes with rising/falling F0). NH Taiwanese outperformed NH Americans across a very large range of sweep rates, and for both tasks (discrimination and labelling), consistent with the idea that speaking a tonal language sharpens the fine-grained internal representation of F0. The largest differences between these populations occurred for the youngest children: NH Taiwanese behaved relatively adult-like at 6 years of age whereas NH Americans progressively caught up with NH Taiwanese up to 18 years of age, presumably compensating their lack of tonal exposure by cognitive skills (more advanced than those of the 6 year olds NH Americans). Given the current slope of regression with chronological age, this tonal language benefit would not be observed in adulthood.

As expected, implanted children struggled in the pitch tasks. Impressively, however, the tonal language benefit transferred to these populations in the labelling task only. This observation opens up a novel account for the tonal language advantage phenomenon that is potentially unrelated to the quality of the internal representation of F0. We speculate that this advantage might partly reflect the familiarity of encountering pitch in a dynamic context, where Taiwanese children avoided confusions with other meanings (e.g. pitch height). Despite stringent limitations in F0 coding, growing up in a tonal language environment could help children in reaching a decision with shallow (for NH) or degraded (for CI) pitch contours.

## Data Availability Statement

All data, materials, and analysis codes are publicly available on the Open Science Framework link: [https://osf.io/7pkjq/?view\\_only=bc343b70162f475a84f135bd2bca8dcf](https://osf.io/7pkjq/?view_only=bc343b70162f475a84f135bd2bca8dcf).

## References

1. Chao, Y.-R. *A Grammar of Spoken Chinese*. University of California Press, Berkeley and Los Angeles (1968).
2. Howie, J. M. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge University Press, Cambridge (1976).
3. Liu, P. & Pell, M. D. Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behav. Res.* **44**, 1042–1051 (2012).
4. Wong, P. C. Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin* **59**, 83–95 (2002).

5. Wong, P. C., Parsons, L. M., Martinez, M. & Diehl, R. L. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J. Neuroscience* **24**, 9153–9160 (2004).
6. Zatorre, R. J. & Gandour, J. T. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical transactions of the royal society of London. Series B, Biological sciences* **363**, 1087–1104 (2008).
7. Krishnan, A. Human frequency following response. In *Auditory evoked potentials: basic principles and clinical application*, edited by Burkard, R. F., Don, M. & Eggermont, J. J., pages 313–335 (2006).
8. Krishnan, A. & Gandour, J. T. The role of the auditory brainstem in processing linguistically relevant pitch patterns. *Brain Lang.* **110**, 135–148 (2009).
9. Xu, Y., Krishnan, A. & Gandour, J. T. Specificity of experience-dependent pitch representation in the brainstem. *Neuro Report* **17**, 1601–1605 (2006).
10. Krishnan, A., Gandour, J. T., Bidelman, G. M. & Swaminathan, J. Experience dependent neural representation of dynamic pitch in the brainstem. *Neuro Report* **20**, 408–413 (2009).
11. Chandrasekaran, B., Krishnan, A. & Gandour, J. T. Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuro Report* **18**, 1963–1967 (2007).
12. Bidelman, G. M., Gandour, J. T. & Krishnan, A. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci* **23**, 425–434 (2011).
13. Stargay, J. R. & Downs, D. Differential sensitivity for frequency among speakers of a tone and a nontone language. *J. Chinese Linguist.* **21**, 143–163 (1993).
14. Bent, T., Bradlow, A. R. & Wright, B. A. The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol.* **32**, 97–103 (2006).
15. Bidelman, G. M., Gandour, J. T. & Krishnan, A. Musicians and tone language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition* **77**, 1–10 (2011).
16. Pfordresher, P. Q. & Brown, S. Enhanced production and perception of musical pitch in tone language speakers. *Attent. Percept. Psychophys.* **71**, 1385–1398 (2009).
17. Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S. & Wicha, N. Y. Y. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Front. Psychol.* **2**, art. 146 (2011).
18. Wong, P. C. M. *et al.* Effects of culture on musical pitch perception. *Plos One* **7**, e33424 (2012).
19. Bidelman, G. M., Hutka, S. & Moreno, S. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *Plos One* **8**, e60676 (2013).
20. Zeng, F. G. Trends in cochlear implants. *Trends Amplif.* **8**, 1–34 (2004).
21. Drullman, R., Festen, J. M. & Plomp, R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **95**, 1053–1064 (1994).
22. Elliott, T. M. & Theunissen, F. E. The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* **5**, e1000302 (2009).
23. Zeng, F. G. Temporal pitch in electric hearing. *Hear. Res.* **174**, 101–106 (2002).
24. Moore, B. C. J. & Carlyon, R. P. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In *Pitch: neural coding and perception*, edited by Plack, C. J., Oxenham, A. J., Fay, R. R. & Popper, A. N. (pages 234–277). New-York Springer/ Birkhauser (2005).
25. Kong, Y.-Y., Deeks, J. M., Axon, P. R. & Carlyon, R. P. Limits of temporal pitch in cochlear implants. *J. Acoust. Soc. Am.* **125**, 1649–1657 (2009).
26. Li, C. N. & Thompson, S. A. The acquisition of tone in Mandarin-speaking children. *J. Child Lang.* **4**, 185–199 (1997).
27. Zhu, H. & Dodd, B. The phonological acquisition of Putonghua (Modern Standard Chinese). *J. Child Lang.* **27**, 3–42 (2000).
28. Peng, S. C. *et al.* Processing of acoustic cues in lexical tone identification by pediatric cochlear implant recipients. *J. Speech Lang. Hear. Res.* **60**, 1223–1235 (2017).
29. Zhou, N., Huang, J., Chen, X. & Xu, L. Relationship between tone perception and production in prelingually deafened children with cochlear implants. *Otology & Neurology* **34**, 499–506 (2013).
30. Deroche, M. L. D., Lu, H.-P., Limb, C. J., Lin, Y.-S. & Chatterjee, M. Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Front. Neurosci.* **8**, 282, <https://doi.org/10.3389/fnins.2014.00282> (2014).
31. Luo, H., Boemio, A., Gordon, M. & Poeppel, D. The perception of FM sweeps by Chinese and English listeners. *Hear. Res.* **224**, 75–83 (2007).
32. Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J. & Chatterjee, M. Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience* **10**, 73, <https://doi.org/10.3389/fnins.2016.00073> (2016).
33. Deroche, M. L. D., Zion, D. J., Schurman, J. R. & Chatterjee, M. Sensitivity of school-aged children to pitch-related cues. *J. Acoust. Soc. Am.* **131**, 2938–2947 (2012).
34. Green, D. M. Psychoacoustics and detection theory. *J. Acoust. Soc. Am.* **32**, 1189–1203 (1960).
35. Macmillan, N. A. & Kaplan, H. L. Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin* **98**, 185–199 (1985).
36. Wichmann, F. A. & Hill, N. J. The psychometric function: I. Fitting, sampling and goodness-of-fit. *Percept. Psychophys.* **63**, 1293–1313 (2001).
37. Wichmann, F. A. & Hill, N. J. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.* **63**, 1314–1329 (2001).
38. Luce, R. D. *Response times*. New York: Oxford University Press (1986).
39. Chatterjee, M. & Peng, S. C. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear. Res.* **235**, 143–56 (2008).
40. Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M. & Wouters, J. Temporal pitch mechanisms in acoustic and electric hearing. *J. Acoust. Soc. Am.* **112**, 621–633 (2002).
41. Carlyon, R. P., Deeks, J. M. & McKay, C. M. The upper limit of temporal pitch for cochlear implant listeners: stimulus duration, conditioner pulses, and the number of electrodes stimulated. *J. Acoust. Soc. Am.* **127**, 1469–1478 (2010).
42. Hong, R. S. & Turner, C. W. Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. *J. Acoust. Soc. Am.* **126**, 291–299 (2009).
43. Gaudrain, E., Deeks, J. M. & Carlyon, R. P. Temporal regularity detection and rate discrimination in cochlear implant listeners. *J. Assoc. Res. Otolaryngology* **18**, 387–397 (2017).
44. Geurts, L. & Wouters, J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J. Acoust. Soc. Am.* **109**, 713–726 (2001).
45. Green, T., Faulkner, A. & Rosen, S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J. Acoust. Soc. Am.* **116**, 2298–2310 (2004).
46. Green, T., Faulkner, A., Rosen, S. & Macherey, O. Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *J. Acoust. Soc. Am.* **118**, 375–385 (2005).
47. Laneau, J., Wouters, J. & Moonen, M. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *J. Acoust. Soc. Am.* **116**, 3606–3619 (2004).
48. Landsberger, D. M., Vermeire, K., Claes, A., van Rompaey, V. & van de Heyning, P. Qualities of single electrode stimulation as a function of rate and place of stimulation with a cochlear implant. *Ear and Hearing* **37**, 149–159 (2016).

49. Fryauf-Bertschy, H. *et al.* Cochlear implant use by prelingually deafened children: the influences of age at implant and length of device use. *J. Speech Lang. Hear. Res.* **40**, 183–199 (1997).
50. Tyler, R. S., Fryauf-Bertschy, H. & Kelsay, D. M. Speech perception by prelingually deaf children using cochlear implants. *Otolaryngol Head Neck Surg* **117**, 180–187 (1997).
51. Nikolopoulos, T. P., O'Donoghue, G. M. & Archbold, S. Age at implantation: its importance in pediatric cochlear implantation. *Laryngoscope* **109**, 595–599 (1999).
52. Kirk, K. I. *et al.* Effects of age at implantation in young children. *Annals of Otolaryngology and Laryngology* **111**, 69–73 (2002).
53. Svirsky, M. A., Teoh, S.-W. & Neuburger, H. Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. *Audiol Neurootol* **9**, 224–233 (2004).
54. Lesinski-Schiedat, A., Illg, A., Heermann, R., Bertram, B. & Lenarz, T. Paediatric cochlear implantation in the first and second year of life: a comparative study. *Cochlear Implants Int.* **5**, 146–154 (2004).
55. Tomblin, J. B., Barker, B. A. & Hubbs, S. Developmental constraints on language development in children with cochlear implants. *Int. J. Audiology* **46**, 512–523 (2007).
56. Dettman, S. J., Pinder, D., Briggs, R. J., Dowell, R. C. & Leigh, J. R. Communication development in children who receive the cochlear implant younger than 12 months: risks versus benefits. *Ear and Hearing* **28**, 11S–18S (2007).
57. Holt, R. F. & Svirsky, M. A. An exploratory look at pediatric cochlear implantation: is earliest always best? *Ear and Hearing* **29**, 492–511 (2008).
58. Houston, D. M., Stewart, J., Moberly, A., Hollich, G. & Miyamoto, R. T. Word learning in deaf children with cochlear implants: effects of early auditory experience. *Develop. Science* **15**, 448–461 (2012).
59. Most, T. & Aviner, C. Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing AIDS, and normal hearing. *J. Deaf Stud. Deaf Educ.* **14**, 449–464 (2009).
60. Chatterjee, M. *et al.* Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear. Res.* **322**, 151–162 (2015).
61. Peng, S. C., Tomblin, J. B. & Turner, C. W. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear.* **29**, 336–351 (2008).
62. Peng, S. C., Lu, N. & Chatterjee, M. Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiol. Neurootol.* **14**, 327–337 (2009).
63. Peng, S. C., Chatterjee, M. & Lu, N. Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends Amplif.* **16**, 67–82 (2012).
64. Nakata, T., Trehub, S. E. & Kanda, Y. Effect of cochlear implants on children's perception and production of speech prosody. *J. Acoust. Soc. Am.* **131**, 1307–1314 (2012).
65. Barry, J. G. *et al.* Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clinical Linguistics and Phonetics* **16**, 79–99 (2002).
66. Ciocca, V., Francis, A. L., Aisha, R. & Wong, L. The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J. Acoust. Soc. Am.* **111**, 2250–2256 (2002).
67. Peng, S. C., Tomblin, J. B., Cheung, H., Lin, Y. S. & Wang, L. S. Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear.* **25**, 251–264 (2004).
68. Gandour, J. T. & Harshman, R. A. Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech* **21**, 1–33 (1978).
69. Francis, A. L., Ciocca, V., Ma, L. & Fenn, K. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *J. Phonetics* **36**, 268–294 (2008).

## Acknowledgements

This work was supported by NIH grants R21 DC011905 and R01 DC014233 awarded to M. Chatterjee, and the Human Subjects Research Core of NIH Grant No. P30 DC004662. We wish to thank all the children for their participation and motivation.

## Author Contributions

M.D. was responsible for experimental design, codes, and data collection at Johns Hopkins. He analysed all the data and wrote the manuscript. H.P.L. collected data in Taiwan. A.K. collected data at the Boys Town site. M.C. and K.B. collected data at the UCSF site. S.C.P. contributed to the rationale of the project and discussion. C.L. and Y.S.L. supervised the project at Johns Hopkins and in Taiwan, respectively. M.C. developed the rationale and obtained funding for this research program. She worked closely with M.D. to design the tasks and edited the manuscript.

## Additional Information

**Competing Interests:** C.L. has received research support from Advanced Bionics (medical advisory board, consultant), Med-El (advisory board), Oticon (consultant), and Spiral Therapeutics (chief medical officer). All other authors declare no competing interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019