CHIN. PHYS. LETT. Vol. 38, No. 2 (2021) 024301

Second Virtual Pitch Shift in Cochlea Observed *In Situ* via Laser Interferometry

Zhang-Cai Long(龙长才)[1†], Yan-Ping Zhang(张艳平)[2†], and Lin Luo(骆琳)[3*]

[1]*School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Affiliated Hospital, Huazhong University of Science and Technology, Wuhan 430074, China*
[3]*Department of Chinese Language and Literature, Huazhong University of Science and Technology, Wuhan 430074, China*

(Received 29 August 2020; accepted 11 December 2020; published online 27 January 2021)

Pitch is the most important auditory perception characteristic of sound with respect to speech intelligibility and music appreciation, and corresponds to a frequency of sound stimulus. However, in some cases, we can perceive virtual pitch, where the corresponding frequency component does not exist in the stimulating sound. This virtual pitch contains a deviation from the de Boer pitch shift formula, which is known as second pitch shift. It has been theoretically suggested that nonlinear dynamics in the cochlea or in the neural network produce a nonlinear resonance with a frequency corresponding to the virtual pitch; however, there is no direct experimental observation to support this theory. The second virtual pitch shift, expressed via basilar membrane nonlinear vibration temporal patterns, and consistent with psychoacoustic experiments, is observed *in situ* in the cochlea via laser interferometry.

DOI: 10.1088/0256-307X/38/2/024301

Pitch is the most important perceptual attribute of sound. A sequence of pitch composes musical melody. The pitch of a vocal sound expresses tone in speech, which is very important with regard to speech intelligibility, particularly in conditions of background noise.[1] For a long time, the underlying mechanism of pitch perception has represented an intriguing, but unresolved scientific problem.[1,2] Persons with serious cochlear hearing loss suffer from a deficit in pitch perception. Neither hearing aids or cochlear implants have succeeded in rehabilitating pitch perception. This makes the recognition of the pitch expressing mechanism in the cochlea, an issue of some significance.[1]

Acoustic pitch is the perceived position of a sound in a musical scale, and can be described in terms of the frequency of a pure tone with the same perceived pitch. Sound with a pitch generally possesses a frequency component equal to the pitch. Hermann von Helmholtz proposed that physical resonance in auditory, with frequency of pitch as the mechanism underlying pitch perception,[3] and Von Bekesy observed this resonance in the cochlea, where different longitudinal positions in the basilar membrane result in different mechanical resonance frequencies. In some cases, however, we perceive pitch where there is no real stimulating component in relation to the frequency of the pitch. For example, when a complex sound, with its fundamental frequency component removed, comprises only higher harmonic components, pitch perception still occurs, and the perceived pitch is that of the removed fundamental component. This pitch, lacking a corresponding physical frequency component, is known as a ghost pitch, or virtual pitch.[4] Von Helmholtz had ever suggested that auditory nonlinearity may produce a frequency component, corresponding to the virtual pitch, which does not exist in the stimulus;[3] these nonlinear products, known as combination tones, were finally observed in cochlear membrane vibration in 1991.[5] According to the combination-tone theory, stimuli with frequencies f_1 and f_2 will produce components with frequencies $nf_1 + mf_2$ (where n and m are integers), and only a complex sound with higher harmonics (i.e., frequencies $2f_0$, $3f_0$,..., nf_0) will produce its missing fundamental frequency component (f_0) (the combination tone with difference frequency of consequent harmonics). However, combination-tone theory is still not the answer. One of the counter evidences to the theory is the phenomenon of virtual pitch shift. When frequencies of higher harmonic components are changed by a common value δf to ($2f_0 + \delta f$, $3f_0 + \delta f$,..., $nf_0 + \delta f$), and the difference frequency of consequent frequencies is f_0 unchanged, perceptual pitch still occurs, and the pitch is not the difference frequency f_0 instead, with a shift from f_0.[6,7] For stimuli with frequencies $\{(k-1)f_0 + \delta f,\ kf_0 + \delta f,\ (k+1)f_0 + \delta f\}$, which have a common change, δf, from harmonics $\{(k-1)f_0,\ kf_0,\ (k+1)f_0\}$, de Boer predicted this pitch shift in formula $f_{\rm pp} = f_0 + \frac{\delta f}{k}$,[5] which is dependent on k and δf. Figure 1 shows typical psychoacoustic measurements of the shift in virtual

Supported by the National Natural Science Foundation of China (Grant Nos. 11374118 and 90820001).
†They contributed equally to this work.
*Corresponding author. Email: luolin@hust.edu.cn
© 2021 Chinese Physical Society and IOP Publishing Ltd

pitch,[6,7] where stimuli were produced via the amplitude modulation of a sinusoidal carrier in the form of $s(t) = \frac{A}{2}[1+\cos(2\pi gt)]\cos[2\pi(kf_0+\delta f)t]$ with k being an integer, which is composed of three components, with frequencies $\{kf_0+\delta f-g,\ kf_0+\delta f,\ f_0+\delta f+g\}$, centered at the carrier frequency $kf_0+\delta f$, and separated by modulating frequency g. The de Boer formula (dashed lines in Fig. 1) is quite consistent with psychoacoustic measurement. However, fastidious scholars have pointed out that there is a slight, yet systematic discrepancy between de Boer's prediction and perceptual pitch. The perceptual data has a slope slightly steeper than the de Boer formula, and the difference in slope increases with k (Fig. 1).[7] The pitch-shift effect of the de Boer prediction is known as the first virtual pitch-shift effect. The slight difference between de Boer's prediction and the psychoacoustic results is referred to as the second virtual pitch-shift effect.

Moreover, there is another phenomenon to be considered in relation to virtual pitch perception: When the carrier frequency is fixed at certain value, and modulating frequency g is increased, the perception virtual pitch decreases with modulating frequency g.[7] This phenomenon is considered to be correlated with, and included in, the second pitch-shift effect.[7]

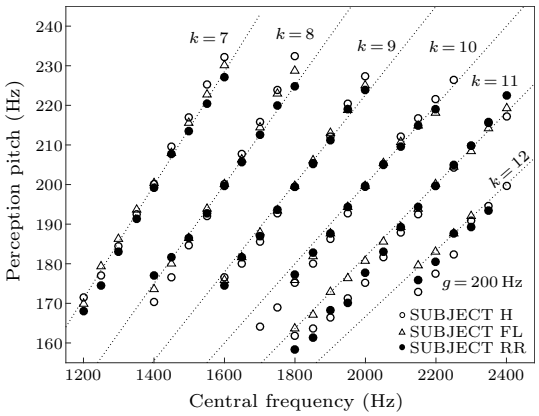

Fig. 1. Perception of pitch as a function of central frequency, measured using Schouten psychoacoustic experiment, from Ref. [7] (dots, circles, and triangles) and predicted by the de Boer formula (dashed line).

To explain the second virtual pitch shift phenomena, a theory involving quasi-periodically forced three-frequency resonance of a nonlinear oscillator has been proposed.[8,9] In this theory, stimuli with frequencies f_1 and f_2, will produce nonlinear resonances with frequency f_R, determined by $Rf_R = pf_1 + qf_2$ (where R, p and q are integers), and the virtual pitch of quasiperiodic sound with shifted harmonic frequencies originates from this nonlinear resonance. For a given stimulus with component frequencies $\{(k-1)f_0+\delta f,\ kf_0+\delta f,\ (k+1)f_0+\delta f\}$, the theoretical virtual pitch is $f_R = f_0 + \frac{\delta f}{k-1/2}$, characterized by a second virtual pitch shift, and a slightly steeper slope than that found using the de Boer formula. It is suggested that this nonlinear resonance for pitch perception may occur in either the cochlea or the neural system. Martignoli and Stoop,[10] using a cascade of Hopf vibrators[11–13] in an electronic device as a cochlear model, demonstrated that second pitch shift phenomena can originate from a temporal structure of auditory peripheral nonlinear vibration. There is, however, no direct physiological experimental evidence for any of these theories. Recently, we examined pitch expression in the cochlea via laser interferometry, and for the first time, in an exsomatized cochlea, observed the first pitch shift phenomena of virtual pitch in relation to basilar membrane vibration; however, there is no second pitch shift, based on the pooled data for the various cochleae used in the research.[14] Where and how the second pitch shift is produced is an issue that remains to be resolved.

In this work, to demonstrate whether second virtual pitch shift exists and originates from nonlinear three-frequency resonance in the cochlea, we measured and analyzed basilar membrane vibration in a cochlea *in situ*, via laser interferometry. A cochlea *in situ* can survive with activity for about three hours,[15] far longer than an exsomatized cochlea, making it possible to obtain more data from an individual cochlea. This enables us to obtain a series of data from an individual cochlea, rather than pooling data from different cochlea in order to compare the subject's pitch perception results.

We used the *in situ* cochlea of guinea pig, which is similar to the human cochlea, to conduct this research. Figure 2 shows a schematic of the research system. The plate-fixed Guinea pig, with cochlea *in situ*, was placed on the platform of a microscope. This microscope was designed by us to introduce measuring laser light from a laser interferometer into the cochlea through its object lens, in such a way that we were able to adjust its platform to locate the measuring point in the cochlea.

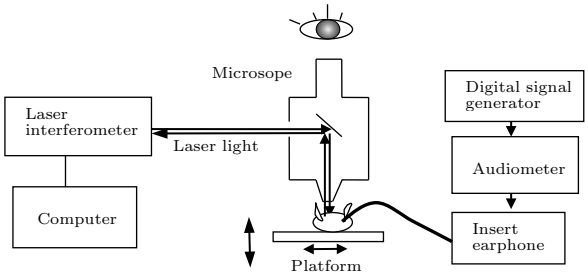

Fig. 2. The schematic of research system.

Stimulating signals were synthesized and produced digitally with 0.01 μs time resolution by a RIGOL DG4062 digital signal generator. These signals took the same form as those used in the psychoacoustic experiments detailed in Refs. [6,7]: $s(t) = \frac{A}{2}[1+\cos(2\pi gt)]\cos[2\pi(kf_0+\delta f)t]$, being composed

of pure tone components with frequencies $\{kf_0+\delta f-g$, $kf_0+\delta f$, $kf_0+\delta f+g\}$. The amplitude ratio of the three components was 0.45 : 1 : 0.45. Signals were introduced to the CD/tape input of an AC33 audiometer (Denmark) and amplified, before being played through an insert earphone (EARTONE-3A) into the guinea pig's ear canal, in order to produce a sound stimulus at an 80 dB sound level.

The laser interferometry measurement system used here was developed by our team. In this system, new technology was adopted to achieve a subnanometer resolution measurement of vibration displacement without nonlinear distortion.[16] In addition, this system was designed to measure the vibration of an object with weak reflection, so that the vibration of the basilar membrane was measured without reflecting micro beads being projected into the cochlea. In this measurement system, the displacement wave forms of basilar membrane vibration, rather than vibration velocity, were recorded in real time, using a computer. Further details regarding this system are available in our published articles[16] and PhD dissertation.[17]

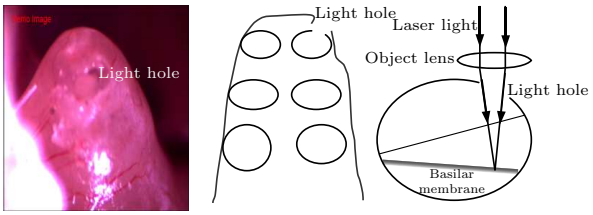

Fig. 3. Measurement position and location in cochlea. Left: cochlea with measurement hole; middle: schematic figure of cochlea with measuring hole; right: location of measurement light in cochlea.

The guinea pigs used here weighed approximately 400 g to 500 g, and were provided by the Experimental Animal Center at the Hubei Provincial Center for Disease Control and Prevention. All Hubei provincial and Chinese national guidelines regarding the care and use of animals were followed. The animals were selected randomly as regards gender and color. Each was examined under the microscope in order to confirm the absence of any abnormality in the ear channel and eardrum. Having been anesthetized via an abdominal injection of pentobarbital sodium (concentration 3%, dose 30 mg/kg), the maxilla, including facial muscle, was removed from each guinea pig to expose its hearing bulla. Under the microscope, the bulla was opened, and a hole measuring about 200 μm was drilled at the top first turn of the cochlea (see the left and middle panels in Fig. 3). Next, the whole guinea pig was placed onto a plate, where the head of the guinea pig was fixed. The plate with cochlea *in situ* was placed onto measuring microscope platform (Fig. 2). By adjusting the height and horizontal position of the platform, the focused measuring laser light through the transparent vestibular membrane was located at the basilar membrane (see the right-hand panel of Fig. 3). The guinea pigs died during the course of the measurement process, due to blood loss and respiratory failure. In order to obtain an active cochlea for the purposes of our research, the available measurement time was therefore limited to less than two hours. In addition, the activity of the cochlea under examination was checked based on whether its basilar membrane vibration contained combination-tone components.[18]

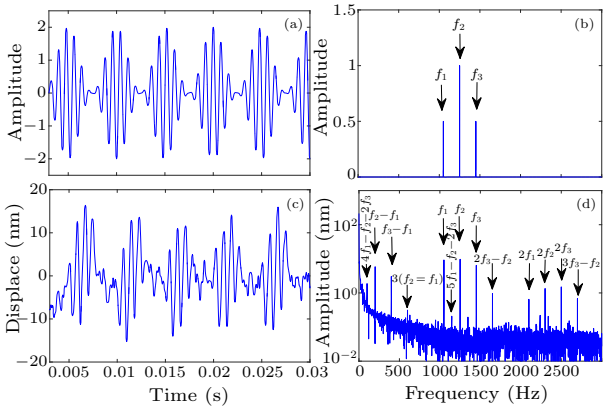

Fig. 4. Stimulating signal and responding vibration of cochlear basilar membrane. (a) Stimulus wave form in the form of $s(t) = \frac{A}{2}[1+\cos(2\pi f_0 t)]\cos[2\pi(kf_0+\delta f)t]$, $f_0 = 200$ Hz, $k = 6$, $\delta f = 50$ Hz. (b) Spectrum of stimulus with component amplitude ratio of 0.45 : 1 : 0.45, $f_1 = 5f_0+\delta f = 1050$ Hz, $f_2 = 6f_0+\delta f = 1250$ Hz, $f_3 = 7f_0+\delta f = 1450$ Hz. (c) Vibration wave form of basilar membrane. (d) Spectrum of basilar membrane vibration, responding to the signal stimulus given in Fig. 2(a), which includes combination tones (200 Hz, 400 Hz, 600 Hz, etc.) in addition to stimulus frequency components ($f_1 = 1050$ Hz, $f_2 = 1250$ Hz, $f_3 = 1450$ Hz). The characteristic frequency of the measurement position was 300 Hz.

Our research aimed to establish whether three-frequency resonance occurs in cochlear membrane vibration. This was achieved by analyzing the frequency components of basilar membrane vibration, and checking whether new frequency components predicted by three-frequency resonance were produced with respect to basilar membrane vibration. Our observation verified the previous results for an exsomatized cochlea, i.e., that no three-frequency resonance appeared in the basilar membrane vibration. Figure 4 depicts a sample observation. It shows that the frequency components [Fig. 4(d)] of the basilar membrane response were simpler than those of the cascade Hopf cochlear model,[10] and that in addition to stimuli components [see Fig. 4(b)] and their combination-tone components, no extraordinary frequency components were predicted in relation to nonlinear three-frequency resonance. Specifically, there is no spectrum peak at the frequency of either hearing perception pitch, the pitch predicted by the de Boer formula, $f_{\rm pp} = f_0 + \frac{\delta f}{k}$, or that predicted by three-frequency resonance, $f_{\rm pp} = f_0 + \frac{\delta f}{k-1/2}$.

Next, we studied the pitch expressed via the tem-

poral structure of the basilar vibration. We obtained the inter-peak time interval distribution of the basilar membrane vibration wave form, and found the most frequent inter-peak time interval, $T_{\rm P}$, of the basilar membrane vibration, for which there is a maximum in the time-interval distribution (see Fig. 3 in Ref. [14]). We use $1/T_{\rm P}$ to denote the temporal structure-expressed pitch, $f_{\rm P}$. This calculation of pitch $f_{\rm P}$ has been adopted in previous theoretical research,[10] and some researchers have suggested that this time information can be maintained in subsequent neuron transfer, and can be extracted in the higher neuronal system.[10]

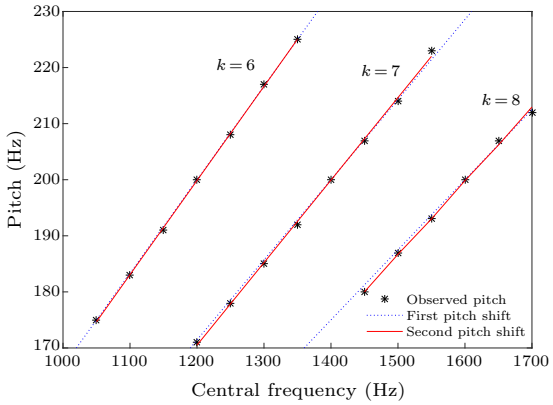

Fig. 5. Pitch shift of basilar membrane vibration. The data in this figure was obtained from an individual cochlea. In this figure, abscissa is the center frequency of stimulus signal $kf_0 + \delta f$, i.e., the carrier frequency of the stimulus signal. The step of δf is 50 Hz; $g = f_0 = 200$ Hz. The slope with dashed lines represents the pitch predicted by the de Boer formula $f_{\rm PP} = f_0 + \delta f/k$. The stars indicate the temporal structure-expressed pitch measured in this research. Solid lines indicate the fit lines for these data. The characteristic frequency of the measuring position was 400 Hz.

Keeping k, f_0 and $g(g = f_0)$ constant, while altering δf, we arrived at a situation in which the three components of stimulus changed around the harmonic. For example, when $k = 6$, $g = f_0 = 200$ Hz, the frequencies of three components, (1000 Hz + δf, 1200 Hz + δf, and 1400 Hz + δf), changed around (1000, 1200 Hz, and 1400 Hz), harmonic frequencies of 200 Hz. We measured the temporal structure-expressed pitch of these stimuli. Owing to the longer active term of the *in situ* cochlea, we were able to obtain data for several different k from a common individual cochlea, rather than pooling data from several cochleae for an individual k, as in our previous research.[14] Figure 3 shows an example of this measurement from a cochlea. It can be seen that when $\delta f = 0$, the temporal-structure expressed pitch is 200 Hz, which is equal to the missing fundamental frequency of stimulus (frequencies 1000 Hz, 1200 Hz, and 1400 Hz), and also the difference frequency of consequent harmonic components of the stimulus. With δf increasing (decreasing), the temporal structure-expressed pitch increases (decreases), and shifts from the original missing fundamental frequency, 200 Hz (this is also the difference frequency). The observed relation of pitch shift with δf (expressed in stars and solid lines as linear fit in Fig. 5) is slightly different in terms of its changing slope from the de Boer formula $f_{\rm pp} = f_0 + \frac{\delta f}{k}$ (see the dashed line slope in Fig. 5). The slope of pitch change with δf, expressed in terms of the temporal structure of the basilar membrane vibration, is slightly steeper than that derived from the de Boer formula, $f_{\rm pp} = f_0 + \frac{\delta f}{k}$, in that it exhibits second pitch shift, as observed in previous psychoacoustic experiments[7] (shown in Fig. 1).

To compare our results with those of previous psychoacoustic experiments, our measurements acquired data corresponding to the psychoacoustic result in Fig. 1. Figure 6 provides an example of these data. It can be observed that the pitch expressed by basilar membrane vibration changed based on the center frequency, in agreement with the psychoacoustic results. A significant characteristic of this result is that the slope is slightly different from, and a little steeper than, that of the de Boer formula, in terms of the so-called second pitch shift. This difference increases with k, as per the results of the psychoacoustic experiments.

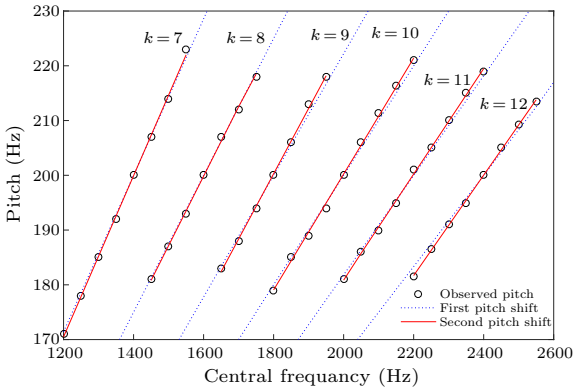

Fig. 6. Temporal structure-expressed pitch, observed in cochlear basilar membrane vibration. The ordinate is the pitch calculated based on the temporal structure of the basilar membrane vibration, abscissa denotes the center frequency of stimulus signal $kf_0 + \delta f$. The stimulus signal form is $s(t) = \frac{A}{2}[1 + \cos(2\pi gt)]\cos[2\pi(kf_0 + \delta f)t]$, $g = f_0 = 200$ Hz. Stars indicate the measured data in this research (solid lines are fit lines). The dashed line slope indicates the pitch described by the de Boer formula. The characteristic frequency of the measuring point is 350 Hz.

Finally, we examined the change of pitch with modulating frequency g. The psychoacoustic measurements show that when the carrying frequency is fixed, perceptional pitch decreases with modulating frequency.[7] This phenomenon, which is considered to represent another occurrence of second pitch shift, cannot be predicted by the de Boer formula. In order to make a comparison with the psychoacoustic measurement, we fixed our carrier frequency at 2000 Hz to measure the change of cochlea-expressed pitch with

modulating frequency, corresponding to the psychoacoustic measurement of pitch in Ref. [7], in which $f_0 = 200\,\text{Hz}$, $k = 10$, and altered the modulating frequency, g, from 190 Hz to 220 Hz. The results are shown in Fig. 7. In the figure, the dashed lines denote the psychoacoustic results of two subjects in Ref. [7], and the cross signs (crosses) indicate the results obtained in this research. It can be seen that the pitch, expressed in cochlear basilar membrane vibration temporal form, changed with modulating frequency, in line with subjective perception pitch.

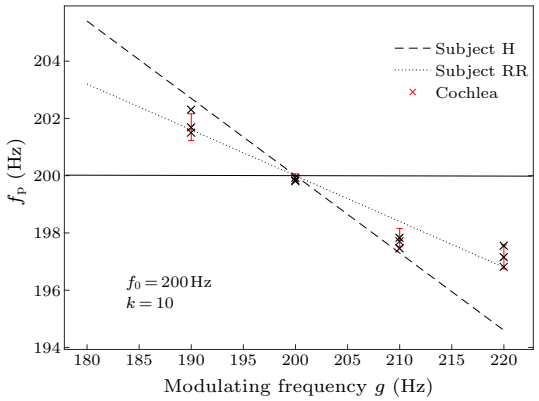

Fig. 7. Pitch change with modulation frequency. Dashed lines denote the subjective results of the psychoacoustic experiment in Ref. [7]. Crosses indicate the cochlear membrane vibration temporal form expressed pitch measured in this research. In these measurements, $f_0 = 200\,\text{Hz}$, $k = 10$, and the characteristic frequency of measurement position of the cochlea basilar membrane was 300 Hz.

To sum up all the above experimental results, it is evident that two aspects of second pitch shift, the pitch-shift deviation from de Boer formula, and the pitch decrease with modulation frequency (space interval of component frequencies), are in agreement with temporal waveform-expressed pitch of cochlear basilar membrane vibration. These characteristics of perception virtual pitch, which the de Boer formula cannot predict, did not require nonlinear three-frequency resonance to be explained, and our experiment has not observed three-frequency resonance phenomenon in relation to cochlear basilar membrane vibration. This indicates that virtual pitch can be produced in the cochlea by nonlinear vibration. However, this nonlinear vibration of the basilar membrane simply produces temporal structure with virtual pitch information, rather than nonlinear resonances, or nonlinear attractors with frequencies corresponding to the virtual pitch. Without a nonlinearity-produced frequency component corresponding to virtual pitch, the expression of virtual pitch in the cochlea is unlike that of the pitch of a single pure tone, in that pitches of tones with different frequencies will map at different positions along the cochlea, due to different resonant characteristic frequencies at different positions in the cochlea, whereas a pure tone pitch is expressed in a very narrow local position. It seems that we cannot expect that the virtual pitch will be expressed at a position in the cochlea at the characteristic frequency of the virtual pitch. In fact, our experiments, which produced results in agreement with those for psychoacoustic perception pitch, did not obtain data at positions in cochlea with the same characteristic frequency as virtual pitch (please refer to the captions in Figs. 5 and 6). Even so, we cannot exclude the possibility that a virtual pitch is expressed in a local area, rather than anywhere in cochlea. The fact that the second pitch shift observed in this work did not appear in the pooled data for different cochlea measured in the previous research means that temporal waveform-expressed pitch may be position-dependent. However, taking into account that the second pitch shift is slight, and that there may be individual differences, we still cannot conclude that virtual pitch expression in the cochlea is position-dependent. Further systematic research into the possible position dependence of virtual pitch would be an important area for future work, which could provide significant scientific input in relation to the development of future cochlear implants, in the form of auditory reconstruction devices with electrodes at different positions in the cochlea to artificially transmit sound information to the brain.

References

[1] Oxenham A J 2013 *Acoust. Sci. Tech.* **34** 388
[2] Moore B C J 2019 *Acoust. Sci. Tech.* **40** 61
[3] von Helmholtz H 1863 *Die Lehre von dem Tonempfindungenals Physiologische Grundlage für die Theorie der Musik* (Braunschweig)
[4] Boer E D 1976 *Handbook of Sensory Physiology: Auditory System* (New York: Springer)
[5] Robles L, Ruggero M A and Rich N C 1991 *Nature* **349** 413
[6] Boer E D 1956 *Nature* **178** 535
[7] Schouten J F, Ritsma R J and Cardozo B L 1962 *J. Acoust. Soc. Am.* **34** 1418
[8] Cartwright J H E, González D L and Piro O 1999 *Phys. Rev. Lett.* **82** 5389
[9] Cartwright J H E, González D L and Piro O 2001 *Proc. Natl. Acad. Sci. USA* **98** 4855
[10] Martignoli S and Stoop R 2010 *Phys. Rev. Lett.* **105** 048101
[11] hEguíluz V M, Ospeck M, Choe Y *et al.* 2000 *Phys. Rev. Lett.* **84** 5232
[12] Tian L, Zhang Y P and Long Z C 2016 *Chin. Phys. Lett.* **33** 128701
[13] Tian L and Long Z C 2017 *Chin. Phys. Lett.* **34** 048702
[14] Long Z C, Zhang Y P and Luo L 2019 *Chin. Phys. Lett.* **36** 024302
[15] Zhang Y P, Huang G, Long X M and Long Z C 2017 *J. Clin. Otorhinolaryngol. Head. Neck Surg.* **31** 1423 (in Chinese)
[16] Wang F, Long Z C, Z B *et al.* 2012 *Rev. Sci. Instrum.* **83** 045112
[17] Wang F 2014 *PhD Dissertation* (Wuhan: Huazhong University of Science and Technology) (in Chinese)
[18] Long X M, Zhang Y P, Lu J and Long Z C 2015 *J. Clin. Otorhinolaryngol. Head. Neck Surg.* **29** 1644 (in Chinese)