

University of Groningen

Frequency selectivity of tonal language native speakers probed by suppression tuning curves of spontaneous otoacoustic emissions

Engler, Sina; de Kleine, Emile; Avan, Paul; van Dijk, Pim

Published in:
Hearing Research

DOI:
[10.1016/j.heares.2020.108100](https://doi.org/10.1016/j.heares.2020.108100)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Engler, S., de Kleine, E., Avan, P., & van Dijk, P. (2020). Frequency selectivity of tonal language native speakers probed by suppression tuning curves of spontaneous otoacoustic emissions. *Hearing Research*, 398, 108100. [108100]. <https://doi.org/10.1016/j.heares.2020.108100>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



ELSEVIER

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Frequency selectivity of tonal language native speakers probed by suppression tuning curves of spontaneous otoacoustic emissions

Sina Engler^{a,b,*}, Emile de Kleine^{a,b}, Paul Avan^{c,d}, Pim van Dijk^{a,b}^a Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, the Netherlands^b Graduate School of Medical Sciences, Research School of Behavioural and Cognitive Neurosciences, University of Groningen, the Netherlands^c Laboratory of Neurosensory Biophysics, University Clermont Auvergne, Laboratory of Neurosensory Biophysics, UMR INSERM 1107, Clermont-Ferrand, France^d School of Medicine, 28 Place Henri Dunant, Clermont-Ferrand 63000, France

ARTICLE INFO

Article history:

Received 12 June 2020

Revised 28 September 2020

Accepted 19 October 2020

Available online 21 October 2020

Keywords:

Auditory

Frequency selectivity

Spontaneous otoacoustic emission

Suppression

Tonal language

ABSTRACT

Native acquisition of a tonal language (TL) is related to enhanced abilities of pitch perception and production, compared to non-tonal language (NTL) native speakers. Moreover, differences in brain responses to both linguistically relevant and non-relevant pitch changes have been described in TL native speakers. It is so far unclear to which extent differences are present at the peripheral processing level of the cochlea. To determine possible differences in cochlear frequency selectivity between Asian TL speakers and Caucasian NTL speakers, suppression tuning curves (STCs) of spontaneous otoacoustic emissions (SOAEs) were examined in both groups. By presenting pure tones, SOAE levels were suppressed and STCs were derived. SOAEs with center frequencies higher than 4.5 kHz were recorded only in female TL native speakers, which correlated with better high-frequency tone detection thresholds. The suppression thresholds at the tip of the STC and filter quality coefficient $Q_{10\text{dB}}$ did not differ significantly between both language groups. Thus, the characteristics of the STCs of SOAEs do not support the presence of differences in peripheral auditory processing between TL and NTL native speakers.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

dB	Decibel
Oct	Octave
SOAE	Spontaneous otoacoustic emission
SPL	Sound pressure level (re): 20 μPa
STC	Suppression tuning curve
f_{SOAE}	SOAE frequency
f_{tip}	STC-tip frequency
TL	Tonal language
NTL	Non-tonal language
PTC	Psychoacoustic tuning curve

Introduction

Languages can be differentiated into tonal (TL) or non-tonal (NTL). Several studies addressed a link between native language and the acuity of pitch perception. In TL, such as Chinese, pitch changes signal different lexical meanings of the same word. There-

fore, in TL, the precise perception of pitch alterations is essential for the understanding of lexical content. It is not surprising that native speakers of TL pay more attention to pitch changes (Braun and Johnson, 2011) and outperform native NTL speakers in pitch interval discrimination (Pfordresher and Brown, 2009; Hove et al., 2010; Giuliano et al., 2011). Producing and perceiving TL-cues may enhance pitch perception (Pfordresher and Brown, 2009; Giuliano et al., 2011) and production (Deutsch et al., 2004). These findings indicate that the individual linguistic background potentially affects pitch perception to some degree.

Depending on native language background, different brain areas become active during a discrimination task of linguistic stimuli (Gandour et al., 1998; 2000). In general, language processing is lateralized to the left-brain hemisphere, whereas tonal pitch processing takes place in the right hemisphere (Zatorre et al., 1994; 2002). Only in TL native speakers, areas of the left brain-hemisphere are activated during pitch processing in a linguistic context (Gandour, 1998). This hemispheric asymmetry might be even expected, as complex linguistic cues are predominantly processed in this hemisphere. When discriminating lexical tones in linguistic contexts, TL native speakers show activation in Broca's area, whereas NTL native speakers do not (Gandour et al., 1998,

* Corresponding author at: Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, the Netherlands

E-mail address: s.engler@umcg.nl (S. Engler).

2000; Wong et al., 2004). Thus, the activation of particular brain areas, depending on the cues of a language and the listener's language experience, indicate specific differences when processing speech.

Language experience may, however, also influence fundamental auditory processing of sounds with no linguistic content (Salmelin et al., 1999; Vihla et al., 2002). For example, an increase in absolute pitch prevalence can be reached by some kind of training effect due to TL acquisition (Deutsch et al., 2004; 2006; Pfordresher and Brown, 2009). In fact, native speakers of TL also have enhanced pitch perception also in non-linguistic contexts, and outperform NTL control groups (Deutsch et al., 2006; Krishnan et al., 2009; Giuliano et al., 2011). Deutsch et al. (2006) described that absolute pitch perception of musically trained TL native speakers is even more precise than that of musically trained subjects with a NTL background. It is unclear, however, whether these enhanced perceptual abilities of TL native speakers are directly related to cochlear frequency selectivity. Fundamental inner ear properties can potentially cause differences in frequency selectivity and pitch perception.

Otoacoustic emissions (initially described by Kemp 1978) allow the non-invasive and objective measurement of frequency selectivity. In the absence of any external stimulation, sounds can be emitted by the ear. These sounds are termed spontaneous otoacoustic emissions (SOAEs). SOAEs are continuous sinusoids with small fluctuations in frequency and level. They can be recorded by placing a sensitive miniature microphone in the ear canal. In humans, otoacoustic emissions are believed to be produced by outer hair cell activity and thus, may indicate healthy hair cell properties (Brownell et al., 1985). The presence of SOAEs is not rare, as approximately 70% of young and normal-hearing humans emit them (Talmadge et al., 1993). SOAEs are, however, not essential for an adequate acoustic perception. Interestingly, human SOAE occurrence differs between genders, with females having a higher SOAE prevalence than males. Moreover, SOAE prevalence differs between ethnic groups, with Asians expressing more SOAEs per ear than Caucasians (Whitehead et al., 1993). What causes these differences in SOAE occurrence remains so far speculative, but could indicate differences in ear properties between Asians and Caucasians.

External tone stimuli have characteristic and frequency selective suppression effects on SOAEs. Suppression tuning curves (STCs) can be derived by measuring the suppression of a single emission peak for various frequencies and levels of the external tone. STCs of SOAEs allow objective and non-invasive estimation of the cochlear frequency selectivity (Schloth and Zwicker, 1983).

The general approach of the current study can be compared to the research of McKetton et al., 2018, who investigated the prevalence of SOAEs and cochlear tuning in participants with and without absolute pitch perception. We examined the cochlear frequency selectivity of Asian subjects with a TL mother tongue and Caucasian subjects with native NTL background, using STCs of SOAEs. We evaluated whether human frequency selectivity at the cochlear level differs systematically between ethnic groups with different native language experience.

Material and methods

Participants

The recordings were carried out in healthy adults, aged between 18 and 31 years. All participants were screened for SOAE occurrence and had normal hearing thresholds at the emission frequency with pure tone thresholds of ≤ 25 dB hearing level (HL). Participants self-reported a clear Asian or Caucasian heritage with either TL or NTL native language experience (respectively). Eight out of 17 participants with TL background and three out of 16 with

NTL background were musically trained. None of them was a professional musician (definition: Micheyl et al., 2006).

This study was approved by the Medical Ethics Committee of the University Medical Center Groningen, Netherlands (Letter of March 11th 2014, METc 2014.099). The Comité d'Evaluation Ethique de l'Inserm (Letter of March 21st 2019, CD/EB 19-034) approved this study in France. The study was conducted in accordance with the Declaration of Helsinki and applicable laws. All participants gave written, informed consent, and received a modest financial compensation for their participation.

Recording procedure

In each participant both ears were examined for the presence of SOAEs. The hearing thresholds of both ears were measured by using pure tone audiometry (Audiosmart, Echodia, Clermont Ferrand, France) at: 0.25, 0.5, 1, 2, 4, and 8 kHz. In each ear in which SOAEs were present, the recording procedure encompassed three main steps: 1) A two-minute SOAE recording without external stimuli. 2) A one-hour suppression measurement, presenting tones over a large number of levels and frequencies, in quasi-random order (exact test duration depended on the number of stimuli presented). 3) A two-minute SOAE recording, equivalent to step 1.

The measurements were conducted at two locations: The University Medical Center Groningen (UMCG, Groningen, Netherlands) and the University Clermont Auvergne (UCA, Clermont Ferrand, France). At the UMCG the measurements were carried out in a double-walled, sound-attenuating chamber (Industrial Acoustics Company, Niederkrüchten, Germany). At the UCA, the measurements were carried out in a quiet office.

SOAE recording

An occluding soft foam ear plug, including the Etymotic ER10-C microphone-speaker system (Etymotic Research, Inc., Elk Grove Village, IL, USA), was placed in the external ear canal. The microphone output was amplified by 40 dB, using the Etymotic ER-10C preamplifier. During the measurements at the UMCG (except for one ear), an additional amplification of 20 dB was applied by using the Stanford Research System (Stanford Research Systems Inc, model SR 640, CA, USA). SOAEs were monitored by feeding the amplified signal into a spectrum analyzer (SRS Inc., model SR 760).

An audiofire AD/DA converter was used to record the microphone signal on the computer disc and to generate the tone stimuli. At the UMCG (Netherlands) we used the Motu 624 (MOTU Inc., MA, USA) for AD/DA conversion, while at the UCA (France) the ESI U 24 XL (ESI Audiotechnik GmbH, Leonberg, Germany) was used. Stimulus generation and response recording was controlled by custom routines developed with Matlab software (MathWorks Inc., 2016a, Natick, MA, USA).

Emission peaks in the time-averaged spectrum that exceeded the noise floor and that were suppressible by external tones were identified as SOAEs. These SOAEs were individual for each ear. The emission recording with the best signal-to-noise ratio was used to calculate the SOAE frequency (f_{SOAE}), the emission width and level. We excluded small frequency components that were not amenable to the Lorentzian curve fit from further analysis. SOAE suppression by at least 3 dB by external tones lower than 70 dB was required as a further inclusion criterion for the STC analysis.

Stimulus presentation

Stimulus tones of different frequencies and levels were presented in a quasi-random order to investigate the suppressive effect of external tones on the SOAEs. The stimulus frequencies were chosen to cover the range in which SOAEs were detected. In most cases, the suppression frequency varied from 500 Hz to 10 kHz in 1/16 octave steps. The stimulus levels ranged from 0 to 70 dB SPL

in 3 dB steps. Thus, the total number of stimuli was 1587. Each stimulus had a duration of 1.2 s (with a 10-ms cosine rise and fall time). For each stimulus tone, a segment of 1.5 s of the microphone signal was recorded and stored. The SOAE recording started 150 ms prior to the tone onset and ended 150 ms after the tone offset. The center one second of this recording was evaluated for suppression effects of the external tone on the SOAE.

Data analysis

To determine the effect of a single external tone on a SOAE, the entire 1-s interval (described above) was analyzed, during which the stimulus tone was present. To include the SOAE, but exclude the stimulus tone and its harmonics, a tonal signal with a frequency equal to the stimulus plus two higher harmonics was fitted to the time-domain of the recorded signal. The resulting fit was subtracted from the recorded signal, creating a residual. A zero-phase band-pass filter with a level response determined by the emission frequency (f_{SOAE}) and the width of the filter (Δf) was applied to the residual to isolate the SOAE of interest:

$$A(f) = \left[1 + (2[f - f_{\text{SOAE}}]/\Delta f)^8\right]^{-1/2} \quad (1)$$

The center frequency of the 60 Hz wide filter was placed at the unsuppressed f_{SOAE} . Subsequently, the Hilbert phase of the filtered signal was used to compute the average emission frequency during the 1-s recording segment. The filter procedure was repeated, with a filter center frequency that now equaled this computed SOAE frequency, and a width narrowed to 10 Hz, for further noise reduction. For stimulus tones closer than this 10 Hz window, SOAE suppression was not assessed. The SOAE level was determined as the average Hilbert envelope during the 1-s interval. This procedure was repeated for each f_{SOAE} and the characteristics of the stimulus, thus creating a full frequency matrix of SOAE levels. Each element of the matrix contained the individual SOAE level for a given stimulus level and -frequency. Sound fragments that contained artefacts (resulting from movements, swallowing, etc.), as determined by an artifact level crossing paradigm, were ignored.

In the further analysis, we only included SOAEs if they were sufficiently strong relative to the noise floor, to create a tuning curve for 3 dB suppression. STCs were characterized by all relevant suppressor-tone frequencies and levels at which the emission reached 3 dB attenuation. Moreover, STCs that consisted of less than 4 data points or were very noisy were excluded from the analysis. The weakest stimulus that produced 3 dB suppression was referred to as the suppression threshold, with a corresponding tip frequency (f_{tip}) of the suppression tuning curve. The STC sharpness was calculated by the filter quality factor $Q_{10\text{dB}}$. This factor is defined as the ratio between f_{tip} and the width of the tuning curve, at 10 dB above the tip ($\Delta f_{10\text{dB}}$):

$$Q_{10\text{dB}} = f_{\text{tip}}/\Delta f_{10\text{dB}} \quad (2)$$

Tuning curves slopes were evaluated for the lower and the higher frequency flank according to the threshold level at f_{tip} (L_1) and 10 dB above threshold (L_2). The corresponding frequencies (f_1 and f_2) were interpolated. The slope S is defined as:

$$S = (L_2 - L_1)/\log_2(f_2/f_1) \quad (3)$$

Results

We included 17 TL native speakers (Chinese) of Asian heritage of whom 13 were female and four were male (see Table 1). The group of NTL native speakers (Dutch, German) consisted of 16 participants of Caucasian heritage; 12 of them were female and four were male. The median age of TL native speakers was 22.2 years, NTL native speakers had a median age of 21.2 years.

Table 1

Overview of evaluated spontaneous otoacoustic emissions (SOAEs) with 3 dB suppression tuning curves (STCs) for both language groups. Number of participants, number of ears producing SOAEs, and number of SOAEs are indicated per group.

Evaluated STCs	Native language group	
	Tonal	Non-tonal
Total N _{Participant}	17	16
Total N _{Ear}	27	23
Total N _{SOAE}	95	86
Female participants		
N _{Participant}	13	12
N _{Ear}	22	18
N _{SOAE}	88	80
Male participants		
N _{Participant}	4	4
N _{Ear}	5	5
N _{SOAE}	7	6

All SOAEs included in this analysis ($n = 181$) showed 3 dB STCs, necessary to evaluate the frequency selectivity. SOAE levels were clearly above the microphone noise. These SOAEs were stable over the time needed to obtain the suppression measurement. In both tested language groups, the number of SOAEs varied between individuals and ears. We recorded 95 SOAEs in 27 ears of Asian TL native speakers and 86 SOAEs in 23 ears of Caucasian NTL native speakers. The majority of the participants ($n = 28$) were tested in the Netherlands. Five female TL native speakers were tested in France.

Spontaneous otoacoustic emissions (SOAEs)

In both native language groups, SOAEs were more often recorded in females than in males (see Table 1). In the TL group, we recorded SOAEs in 22 ears of females and in five ears of males. In the NTL group, we recorded SOAEs in 18 ears of females and five ears of male participants. Moreover, females did not only tend to have SOAEs more commonly, they also had more emissions per emitting ear. In native speakers of TL, SOAEs of females represent 93% of all recorded SOAEs ($n = 95$). In the NTL group, also 93% of all recorded SOAEs ($n = 86$) were recorded in females.

The frequency distribution of the SOAEs is shown in Fig. 1, and was similar between both language groups (Kolmogorov-Smirnov test, $p = 0.176$). TL native speakers had SOAEs between 0.63 and 8.53 kHz (median: 1.84 kHz). In the NTL native speakers, SOAEs ranged from 0.60 kHz to 4.47 kHz (median: 1.84 kHz). Thus, the SOAEs of TL native speakers were recorded in a wider frequency range towards the higher frequencies. In seven ears (26%) of TL native speakers, SOAEs with frequencies larger than 4.5 kHz were recorded. The hearing sensitivity of these ears was not exceptionally good at these frequencies. However, in general, TL native speakers had relatively good hearing thresholds over the frequency range from 2 to 8 kHz with mean audiometric thresholds between 0.4 and 2.2 dB HL (Fig. 1A). An independent sample t -test with Bonferroni correction revealed significant hearing threshold differences between both language groups at 0.5 kHz ($p < 0.001$), 1 kHz ($p = 0.002$), and 8 kHz ($p < 0.001$). Both language groups had similar SOAE levels (Kolmogorov-Smirnov test, $p = 0.251$), as can be seen in Fig. 2A. Both language groups show differences in emission levels across frequencies that were related to the higher f_{SOAEs} recorded in the TL group. In TL native speakers, no correlation between the SOAE level and f_{SOAE} was found ($R^2 = 0.02$; $p = 0.18$), also when evaluating f_{SOAEs} up to 4.5 kHz ($R^2 = 0.02$; $p = 0.21$) only. In the NTL group, however, a weak negative correlation between emission level and frequency was found ($R^2 = 0.27$; $p < 0.000$). In both language groups the SOAE width was negatively correlated

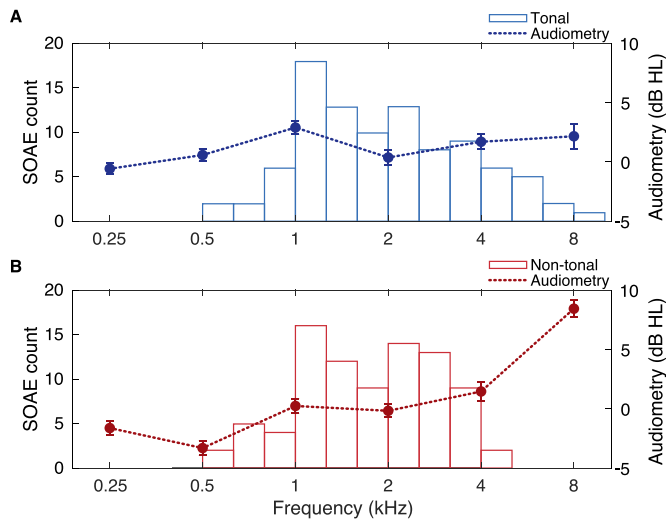


Fig. 1. Distribution of SOAEs and hearing thresholds. Panels show the number of SOAEs per frequency band and the corresponding average pure tone thresholds, per group. Each SOAE count corresponds to one emission peak from which the STC was obtained ($n = 181$). (A) SOAEs of TL native speakers covered a frequency range from 0.63 to 8.52 kHz ($n = 95$). Panel (B) shows the SOAEs of NTL native speakers ($n = 86$). SOAEs of this language group cover a range from 0.60 to 4.46 kHz. Mean pure tone thresholds are plotted with standard error bars. Significant differences in hearing threshold between both language groups were present at 0.5, 1, and 8 kHz.

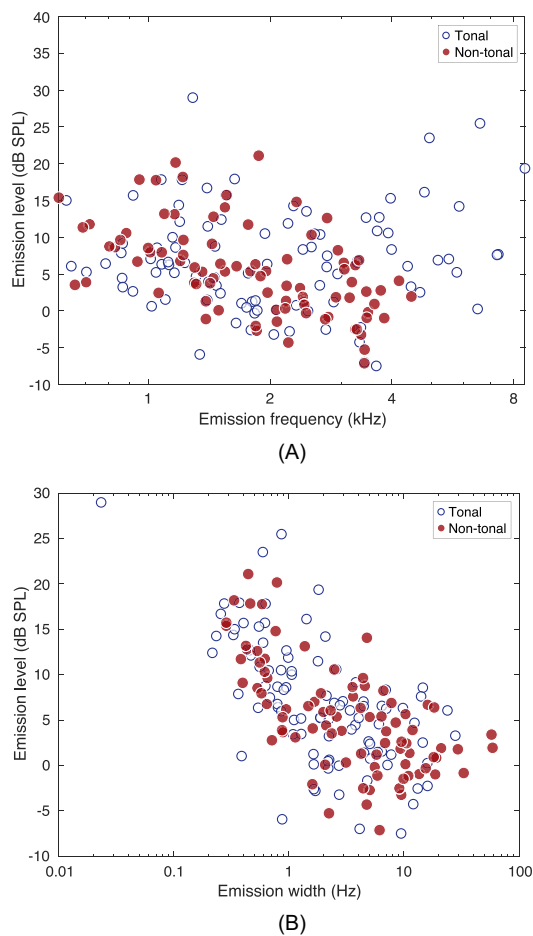


Fig. 2. Characteristics of unsuppressed SOAEs. In both panels open circles indicate data of TL and filled circles that of NTL native speakers. (A) The relationship between SOAE frequency and level. (B) The relationship between SOAE level and width. In both language groups the SOAE width was significantly negatively correlated with the SOAE level (TL: $p < 0.001$, $R^2 = 0.13$ and NTL: $p = 0.001$, $R^2 = 0.12$).

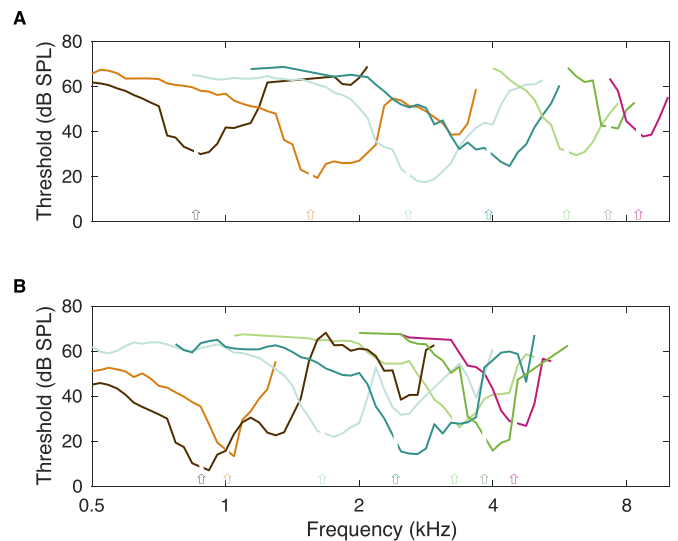


Fig. 3. Representative STCs of SOAEs from different TL (A) and NTL native speakers (B). STCs indicate the stimulus level needed for 3 dB suppression of the SOAE. The arrows indicate the SOAE frequencies. The colors match the corresponding STC. The stimulus frequencies within 10 Hz of the unsuppressed SOAE frequency were omitted (see main text) and appear as gaps in the STC. Note that in both language groups some STCs contain secondary suppression lobes.

with the SOAE level (Fig. 2B). Thus, large emission peaks were significantly narrower than SOAEs with smaller levels (using ANOVA) in the TL ($R^2 = 0.13$; $p < 0.001$) and NTL ($R^2 = 0.12$; $p = 0.001$) group.

In summary, for both language groups, we found the well-known difference that females show more SOAEs than males. Compared to the NTL native speakers, TL native speakers had better hearing thresholds and more SOAEs at frequencies between 4.5 and 8 kHz.

Suppression tuning curves (STCs)

Here, STCs of TL and NTL native speakers were compared, to evaluate whether increased frequency selectivity could also be observed at cochlear level. STCs were asymmetrically V-shaped and selectively tuned (Fig. 3). We evaluated the slopes of both STC flanks of the TL native speakers ($n = 70$) and the NTL native speakers ($n = 69$). Average STC slopes did not differ significantly between groups. The average low frequency slopes were 35 and 41 dB/oct, and the average high frequency slopes were 46 and 45 dB/oct, respectively, for the TL and NTL group.

The f_{tip} (most sensitive frequency) of the STC could fall on either side of the emission frequency, but was typically above the emission frequency. The f_{tip} in TL native speakers was on average 5.8% above the unsuppressed SOAE frequency, versus 4.3% in the NTL group. The level at f_{tip} (STC's best threshold) between both native language groups differed in their median with 3 dB, median thresholds being 22.1 dB SPL for the TL native speakers and 19.0 dB SPL for the NTL native speakers. We were interested whether SOAE level correlates with suppression threshold (Fig. 4). Interestingly, the suppression thresholds in TL native speakers were independent from SOAE level, whereas the suppression threshold was significantly negatively correlated to emission level in the NTL group ($p < 0.001$). When evaluating the frequency range up to 4.5 kHz only, the suppression threshold remained independent from the SOAE level in the TL group.

Tuning curve sharpness and tuning curve side-lobes

STCs showed the typical asymmetric shape, with steeper slopes for the high-frequency flank. In Fig. 5A we show the average STC

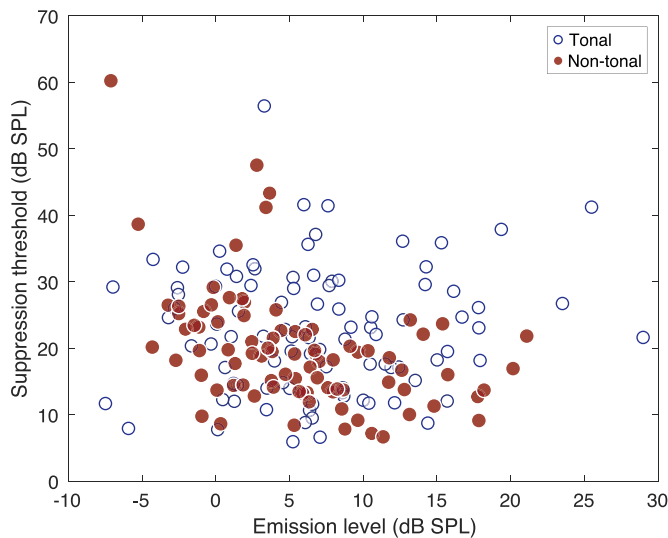


Fig. 4. The suppression thresholds of 3 dB STCs in relation to SOAE level. Open circles indicate data of TL and filled circles that of NTL native speakers. In the TL group the suppression threshold was not correlated with emission level, whereas in the NTL group both were significantly ($p < 0.001$) correlated.

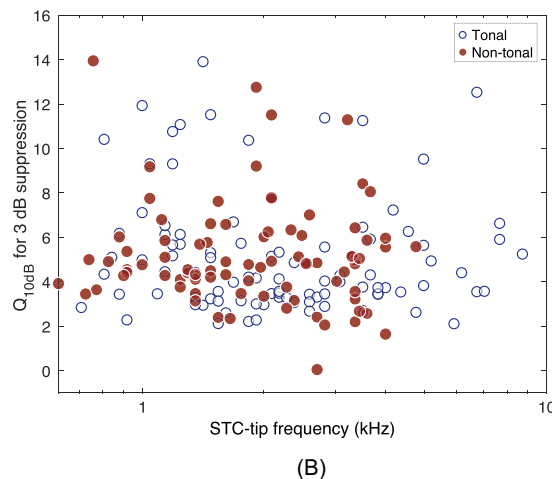
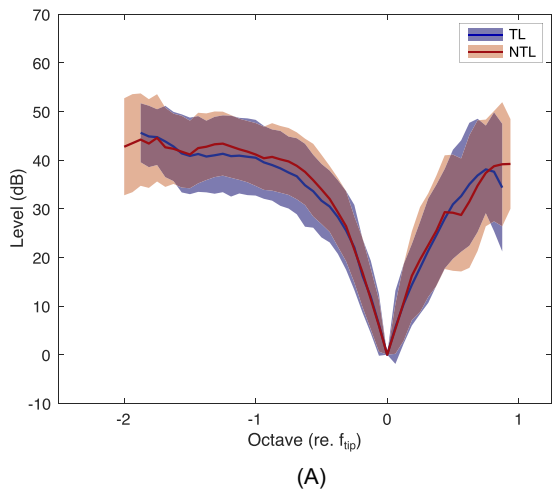


Fig. 5. Comparison of frequency selectivity between both language groups. (A) The average STC per language group with standard deviation (sd). Data shown represents at least 10% of the STCs. (B) Comparison between filter quality measures between TL (open symbols) and NTL native speakers (filled symbols). The filter quality factor Q_{10dB} was determined from STCs. Tuning selectivity in both language groups was similar and did not correlate with STC-tip frequency.

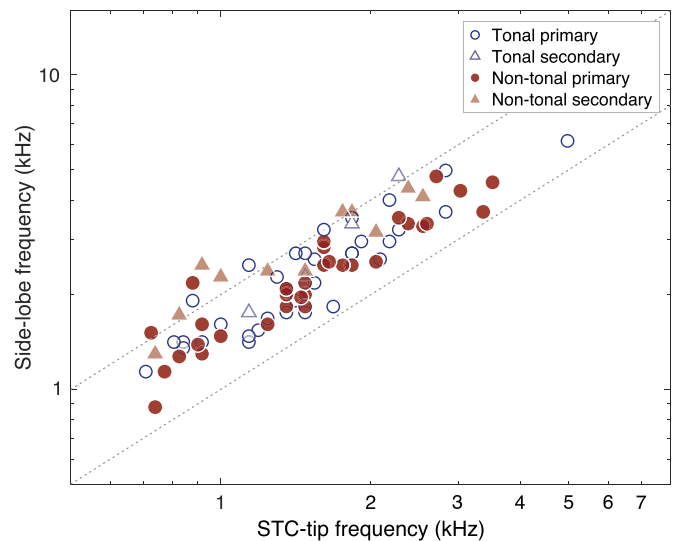


Fig. 6. Frequencies of the primary and secondary side-lobes of STCs, as a function of the frequency of the primary STC-tip. The dashed diagonal lines are added for orientation and are 1 octave apart. Values of the side-lobes of TL (open circles) and NTL native speakers (filled circles) and secondary side-lobes of TL (open triangles) and NTL (filled triangles) are shown. In both language groups the average ratio between the STC-tip and the first side-lobe is 1:1.5 (sd TL: 0.3; sd NTL: 0.3). The average ratio between the STC-tip and the second side-lobe is in the TL group 1:1.8 (sd: 0.3) and in the NTL group 1:2.0 (sd: 0.3).

per language group, that represents at least 10% of the data, with standard deviation. All STCs were aligned with respect to the tip frequency and level. The averaged STCs were very similar between both language groups. The frequency selectivity was defined as the filter quality factor Q_{10dB} , for all subjects (Fig. 5B). Tuning was similar in both language groups (median Q_{10dB} TL: 4.28; median Q_{10dB} NTL: 4.81) and independent from f_{tip} .

Side-lobes represent an additional suppression area at the higher frequency flank of the STC, in some cases even two side-lobes could be observed (Fig. 3). In Fig. 6 we show the frequency of the side-lobes, as a function of frequency of the main STC-tip. Primary side-lobes were in general 0.5–1 octave above the emission frequency. We observed STCs with primary side-lobes in 38% of the emissions for the TL group and in 37% of the NTL group. Additional secondary side-lobes were recorded less commonly. Of the STCs with primary side-lobes, the TL native speakers rarely had secondary side-lobes (11%), whereas secondary side-lobes in the NTL group were recorded more frequently (38%).

Discussion

The properties of spontaneous otoacoustic emissions (SOAEs) were compared between native speakers of a tonal language (TL) and those of a non-tonal language (NTL). SOAEs of both language groups were similar in all aspects we investigated, except for the range of frequencies at which SOAE were detected. In the TL group, emissions were detected above 4.5 kHz, whereas SOAE frequencies in the NTL group never exceeded this frequency. We specifically evaluated frequency tuning curves of SOAE suppression, because of the possible role of cochlear frequency selectivity in language processing. However, we found no difference in frequency selectivity between both language groups.

Our findings correspond with previous research that reported that Asians are more likely to emit SOAEs at higher frequencies compared to Caucasians (Whitehead et al., 1993; Chan and McPherson, 2001). The occurrence of high frequency SOAEs can potentially be caused by (1) middle ear and (2) inner ear differences between both groups. In general, a shorter meatus, smaller mid-

dle ear canal volume, and a smaller tympanic membrane, increase the high frequency transmission (Plassmann and Brändle, 1992). Models have also shown that such changes in middle ear mechanics influence otoacoustic emissions (Avan et al., 2000). On average, Asians have smaller ear canal volumes when compared to Caucasian subjects (Whitehead et al., 1993; Chan and McPherson, 2001; Wan and Wong, 2002; Shahnaz and Davies, 2006). Asians who emitted SOAEs at higher frequencies in fact had smaller ear canal volumes and static admittance than Caucasians (Chan and McPherson, 2001). Consequently, the middle ear characteristics of Asians may favor the transmission of high frequencies. This would not only affect the SOAE transmission towards the outside but also the transmission of high frequency sounds into the ear, which could explain lower hearing thresholds at higher frequencies in TL native speakers (Fig. 1A).

Recently, peripheral frequency selectivity has been investigated using a number of measures. All these measures provide an estimate of the quality factor of cochlear filters, either expressed as Q_{10dB} or Q_{ERB} . In general, measures which are believed to be unaffected by cochlear nonlinearity (compression), provide relatively high values for the quality factor of cochlear filters. In humans, Q-values obtained using these methods range from about 15 to 20, where larger Q-values are measured for cochlear filters with higher center frequency. These linear methods include measurements of stimulus-frequency otoacoustic emissions (SFOAE) group delays (Shera et al., 2002) and forward masking psychoacoustic tuning curves (Shera et al., 2002; Oxenham and Shera, 2003). In contrast, measurements of peripheral frequency selectivity, that depend on cochlear compression, provide broader filter estimates with Q-values of approximately 4–5. These methods include measurements of suppression tuning curves of otoacoustic emissions (SOAEs: Zizz and Glattke, 1988; Manley and van Dijk, 2016; SFOAEs: Charaziak and Siegel, 2014; DPOAEs: Abdala et al., 2007) and psychoacoustic tuning curves derived by simultaneous masking (Moore, 1978; Oxenham and Shera, 2003). Smaller Q-values are presumably based on the compressive action of the mechanical response of the basilar membrane, which shows broader spatial excitation patterns at higher sound levels (Robles and Ruggero, 2001).

Notably, SOAE suppression, as studied in the current paper, must inherently depend on cochlear compression. In general, when two signals are processed by a compressive nonlinearity, the stronger signal determines the degree of compression, which then affects the smaller signal more than if it was present alone. The smaller signal is thus suppressed. Hence, when the external tone interacts with the SOAE, the latter will be suppressed when the excitation due to the tone becomes larger than the excitation of the SOAE. Models suggest that the vibration pattern of a SOAE is maximal near the tonotopic place corresponding to the emission frequency (Epp et al., 2015; for an animation see the supplemental material of Manley and van Dijk, 2016). Consequently, it can be assumed that SOAE suppression by an external tone reflects tonotopy and frequency selectivity of the basilar membrane.

In short, measures of cochlear tuning that are believed to reflect near-threshold linear cochlear behavior, provide highly selective estimates of cochlear tuning. There is substantial evidence that the Q-values of these linear measurements correspond to those of auditory neurons (Shera et al., 2002; Joris et al., 2011; Sumner et al., 2018), although the comparison includes the assumption of a factor, referred to as the "tuning ratio". In contrast, in measurements based on experimental protocols that presumably engage cochlear compression, estimates of Q-values are lower. These nonlinear measurements suggest broader tuning, which reflects the fact that cochlear mechanical excitation patterns are wider at higher sound levels. Consistently, the Q-values of SOAE suppression tuning are below that of neural tuning curves in mammals (macaque: cf. Martin et al., 1988 with Joris et al., 2011) and also birds (barn

owl: cf. Engler et al., 2020 with Köppl, 1997a, Köppl, 1997b). In the present study, we used the nonlinear measurement paradigm of SOAE suppression tuning. As described above, this measurement presumably reflects cochlear frequency selectivity, although it may not be a direct measure of nerve fiber tuning. Nevertheless, differences in mechanical cochlear tuning between TL and NTL participants likely would have been detected if they existed. Hence, our results suggest that there is no difference in tuning of the basilar membrane between TL and NTL speakers.

The only aspect where STC were different between TL and NTL speakers was the number of secondary side-lobes that were observed. In TL speakers, these side-lobes were more common than in NTL speakers. At present we can only speculate about an explanation for this difference. SOAEs are believed to correspond to standing waves in the cochlea (Kemp, 1980; Shera, 2003; Epp et al., 2015). In a model of basilar membrane mechanics, these standing waves have antinodes at well-defined positions along the basilar membrane (Epp et al., 2015). Manley and Van Dijk (2016) suggested that the tonotopic frequency of these antinode positions corresponds with side-lobes in the STC. Thus, the side-lobes may be a consequence of interactions between the external tones in the STC measurements and the antinodes of the standing wave. Note, that the standing wave occurs between the stapes footplate of the middle ear and the tonotopic location of the SOAE frequency. Possibly, the differences between side-lobe properties found in the current study may reflect subtle differences in the mechanical properties of the middle ear, as described earlier in this section. However, at present this remains speculative.

Behavioral studies have shown that Asian TL native speakers outperform Caucasian NTL native speakers in pitch perception accuracy (e.g.: Deutsch et al., 2004; Pfordresher and Brown, 2009; Hove et al., 2010; Braun and Johnson, 2011; Giuliano et al., 2011). Our aim was to evaluate whether this enhanced pitch perception of TL speakers reflects sharper frequency selectivity at cochlear level. As a measure of cochlear frequency selectivity, we evaluated STCs of SOAEs. Possible mechanisms behind the enhanced pitch perception of TL native speakers, and to what extent it has a basis in cochlear frequency selectivity, is discussed in the following paragraphs.

The absence of a difference in cochlear tuning suggests that more central structures are responsible for the better pitch acuity in TL speakers. Speech processing is mainly mediated in the left-brain hemisphere, the area where also the temporal information is encoded (Zatorre et al., 2002). Studies have shown that human speech understanding is primarily achieved by temporal processing rather than frequency selectivity (Shannon et al., 1995). For TL native speakers the detection of time varying pitch contours is essential for their native language understanding. At the cochlear level, wider auditory filters would theoretically lead to an improvement of temporal processing. However, wider auditory filters would cause poorer spatial resolution which consequently results in a decrease of frequency selectivity. We did not detect such a significant difference in Q_{10dB} for TL native speakers. Therefore, there was neither evidence for enhanced frequency selectivity nor for better temporal processing at cochlear level of Asian TL speakers, that would explain their behavioral outperformance in pitch perception.

Acoustical training is linked to the development of enhanced pitch perception. This training effect appears to generalize across linguistic and non-linguistic specific contexts. Musicians, for example, do not only detect frequency-movements of pure tones very precisely, but also perceive pitch-contours in linguistic manipulations more accurately (e.g.: musician children: Magne et al., 2006; professional musicians: Schön et al., 2004). Therefore, musical training seems to favor the processing of linguistic relevant pitch information in Mandarin Chinese (Wong et al., 2007). Moreover, TL

acquisition (as a form of acoustical training) leads to pitch accuracy in non-linguistic contexts as well (e.g. Salmelin et al., 1999; Vihla et al., 2002; Deutsch et al., 2004; 2006; Pfordresher and Brown, 2009). In this study we included non-professional musicians only, thus any possible training effect would be primarily related to the native language and therefore stable within each group.

Acoustical experience favors the accurate behavioral perception of tones, as the TL acquisition is linked to an increased accuracy of tone perception (Deutsch et al., 2004; 2006; Pfordresher and Brown, 2009). We speculate that acoustical training, due to language acquisition, enhances the pitch perception abilities of TL native speakers. Studies have addressed a link between TL experience and enhanced pitch representation and tracking. However, this enhancement could not be fully explained by increased temporal pitch processing (Krishnan et al., 2005). Therefore, it was hypothesized that language experience induces neural plasticity at the brainstem level. In fact, TL native speakers showed enhanced pitch encoding measured at the brainstem (Krishnan et al., 2005) and cortical pathways (Kuhl, 2004). In other words, language experience affects the neural pathways at subcortical brainstem level and the central level of the auditory cortex.

Moreover, behavioral and imaging studies have shown that speech processing networks develop which are language-specific (e.g. Gandour et al., 2000; Zatorre et al., 2002). When infants learn their native language, their brains develop language-specific networks (Kuhl, 2004; Krishnan et al., 2010b). TL native speakers seem to use different neural networks depending on whether the change in pitch is linguistically relevant or not (Gandour et al., 1998; Wong et al., 2004; Pfordresher and Brown, 2009). Thus, the sensitivity is increased to sounds that are similar to those of this particular language (Kraus and Banai, 2007; Krishnan et al., 2010a). Consequently, the auditory system appears to be experience-based and plastic in modification. Experience dependent neural ascending and descending pathways optimize the functionality and form the auditory cortex (Suga et al., 2003). Interestingly, such processing pathways are not strictly restricted to language-specific cues (e.g. Bent et al., 2006). TL acquisition tunes the overall neuronal response to pitch in the brainstem with enhanced sensitivity to speech relevant cues (e.g.: Swaminathan et al., 2008; Krishnan et al., 2009). Thus, effects of acoustical training can generalize across linguistic and non-linguistic specific contexts (e.g.: Bidelman et al., 2013). Presumably, this is how TL acquisition (as a form of acoustical training) can lead to pitch accuracy in non-linguistic context as well (e.g. Salmelin et al., 1999; Vihla et al., 2002; Deutsch et al., 2004; 2006; Bent et al., 2006; Pfordresher and Brown, 2009).

In addition to differences in acoustical training and exposure, there are also differences in gene expression associated with pitch perception. Specific genes may be linked to enhanced pitch perception (Zatorre, 2003; Schellenberg and Trehub, 2008; Hove et al., 2010). This aspect becomes especially important when testing whether native language experience can be ruled out as a training factor for pitch perception, for example when testing Asians that grew up without exposure to a TL.

Conclusions

In this study, SOAEs of Asians with TL acquisition and Caucasians with no TL experience were recorded and suppressed by pure-tone stimulation. Suppression tuning curves were similar between both language groups. This suggests that the enhanced frequency selectivity of Asian TL native speakers is not based on a difference in cochlear processing. SOAEs above 4.5 kHz were found in TL native speakers only, which is probably based on differences in middle ear properties.

Contributors

S.E., P.v.D., and E.d.K. designed the study. S.E. performed the measurement, wrote the manuscript and visualized the data. All authors, S.E., P.v.D., E.d.K., and P.A., conducted the analysis then verified and approved the final manuscript.

Acknowledgments

This work was supported by the EGRET (European Glaucoma Research Training Program, Marie Skłodowska-Curie-COFUND grant; No 661883). A special thanks to Paolo Toffanin for programming.

References

- Abdala, C., Keefe, D.H., Oba, S.I., 2007. Distortion product otoacoustic emission suppression tuning and acoustic admittance in human infants: birth through 6 months. *J. Acoust. Soc. Am.* 121 (6), 3617–3627.
- Avan, P., Büki, B., Maat, B., Dordain, M., Wit, H.P., 2000. Middle ear influence on otoacoustic emissions. I: noninvasive investigation of the human transmission apparatus and comparison with model results. *Hear. Res.* 120, 189–201.
- Bent, T., Bradlow, A.R., Wright, B.A., 2006. The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum.* 32 (1), 97–103.
- Bidelman, G.M., Hutka, S., Moreno, S., 2013. 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 (4), 1–11.
- Braun, B., Johnson, E.K., 2011. Question or tone 2? How language experience and linguistic function guide pitch processing. *J. Phon.* 39, 585–594.
- Brownell, W.E., Bader, C.R., Bertrand, D., de Ribapierre, Y., 1985. Evoked mechanical responses of isolated cochlear outer hair cells. *Science* 227 (4683), 194–196.
- Chan, J.C.Y., McPherson, B., 2001. Spontaneous and transient evoked otoacoustic emissions: a racial comparison. *JAM* 10, 20–32.
- Charaziak, K.K., Siegel, J.H., 2014. Estimating cochlear frequency selectivity with stimulus-frequency otoacoustic emissions in chinchillas. *J. Assoc. Res. Otolaryngol.* 15, 883–896.
- Deutsch, D., Henthorn, T., Dolson, M., 2004. Absolute pitch, speech, and tone language: some experiments and a proposed framework. *Music Percept.* 21 (3), 339–356.
- Deutsch, D., Henthorn, T., Marvin, E., Xu, H.S., 2006. Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119 (2), 719–722.
- Epp, B., Hero, W., van Dijk, P., 2015. Clustering of cochlear oscillations in frequency plateaus as a tool to investigate SOAE generation. In: Karavitaki, K.D., Corey, D.P. (Eds.). *Mechanics of Hearing: Protein to Perception, Proceedings of the 12th International Workshop on the Mechanics of Hearing, 1703*. AIP Conf. Proc 090025-1-090025-6.
- Engler, S., Köppl, C., Manley, G.A., de Kleine, E., van Dijk, P., 2020. Suppression tuning of spontaneous otoacoustic emissions in the barn owl (*Tyto alba*). *Hear. Res.* 385, 1–9.
- Gandour, J., 1998. Aphasia in tone languages. In: Coppens, P., Basso, A., Lebrun, Y. (Eds.), *Aphasia in Atypical Populations*. Lawrence Erlbaum, Hillsdale, NJ, pp. 117–141.
- Gandour, J., Wong, D., Hutchins, G.D., 1998. Pitch processing in the human brain is influenced by language experience. *Neuroreport* 9 (2), 2115–2119.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., van Lancker, D., Hutchins, G.D., 2000. A crosslinguistic PET study of tone perception. *J. Cogn. Neurosci.* 12 (1), 207–222.
- Giuliano, R.J., Pfordresher, P.Q., Stanley, E.M., Narayana, S., Wicha, N.Y.Y., 2011. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Front. Psychol.* 146 (2), 1–12.
- Hove, M.J., Sutherland, M.E., Krumbhansl, C.L., 2010. Ethnicity effects in relative pitch learning. *Psychon. Soc., Inc.* 17 (3), 310–316.
- Joris, P.X., Bergevin, C., Kalluri, R., Mc Laughlin, M., Michelet, P., van der Heijden, M., SHERA, C.A., 2011. Frequency selectivity in old-world monkeys corroborates sharp cochlear tuning in humans. *Proc. Nat. Acad. Sci.* 108 (42), 17516–17520.
- Kemp, D.T., 1978. Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* 64 (5), 1386–1391.
- Kemp, D.T., 1980. Towards a model for the origin of cochlear echoes. *Hear. Res.* 2, 533–548.
- Köppl, C., 1997a. Frequency tuning and spontaneous activity in the auditory nerve and cochlear nucleus magnocellularis of the barn owl *Tyto alba*. *J. Neurophysiol.* 77, 364–377.
- Köppl, C., 1997b. Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*. *J. Neurosci.* 17 (9), 3312–3321.
- Kraus, N., Banai, K., 2007. Auditory-processing malleability. *Focus on language and music. Assoc. Physiol. Sci.* 16 (2), 105–110.
- Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168.

- Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang.* 110 (3), 135–148.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010a. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Res.* 8 (1313), 124–133.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010b. The effects of tone language experience on pitch processing in the brainstem. *J. Neurolinguist.* 23 (1), 81–95.
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.
- Magne, C., Schön, D., Besson, M., 2006. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18 (2), 199–211.
- Manley, G.A., van Dijk, P., 2016. Frequency selectivity of the human cochlea. *Hear. Res.* 336, 53–62.
- Martin, G.K., Lonsbury-Martin, B.L., Probst, R., Coats, C.A., 1988. Spontaneous otoacoustic emissions in a nonhuman primate. I. Basic features and relations to other emissions. *Hear. Res.* 33, 49–68.
- McKetton, L., Purcell, D., Stone, V., Grahm, J., Bergevin, C., 2018. No otoacoustic evidence for a peripheral basis of absolute pitch. *Her. Res.* 370, 201–208.
- Moore, B.C.J., 1978. Psychophysical tuning curves measured in simultaneous and forward masking. *J. Acoust. Soci. Am.* 63, 524–532.
- Oxenham, A.J., Shera, C.A., 2003. Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J. Assoc. Res. Otolaryngol.* 4, 541–554.
- Pfordresher, P.Q., Brown, S., 2009. Enhanced production and perception of musical pitch in tone language speakers. *Attent. Percept. Psychophys.* 71 (6), 1385–1398.
- Plassmann, W., Brändle, K., 1992. A functional model of the auditory system in mammals and its evolutionary implications. In: Webster, D.B., Fay, R.R., Popper, A.N. (Eds.), *The Evolutionary Biology of Hearing*. Springer-Verlag, New York, pp. 637–653.
- Robles, L., Ruggero, M.A., 2001. Mechanics of the mammalian cochlea. *Physiol. Rev.* 81 (3), 1305–1352.
- Salmelin, R., Schnitzler, A., Parkkonen, L., Biermann, K., Helenius, P., Kiviniemi, K., Kuukka, K., Schmitz, F., Freund, H., 1999. Native language, gender, and functional organization of the auditory cortex. *Proc. Natl. Acad. Sci.* 96 10466–10465.
- Schellenberg, E.G., Trehub, S.E., 2008. Is there an Asian advantage for pitch memory? *Music Percept.* 25 (3), 241–252.
- Schloth, E., Zwicker, E., 1983. Mechanical and acoustical influences on spontaneous oto-acoustic emissions. *Hear. Res.* 11, 285–293.
- Schön, D., Magne, C., Besson, M., 2004. The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Shahnaz, N., Davies, D., 2006. Standard and multifrequently tympanometric norms for Caucasian and Chinese young adults. *Ear Hear.* 75–90.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270 (5234), 303–304.
- Shera, C.A., Guinan, J.J., Oxenham, A.J., 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci.* 99 (5), 3318–3323.
- Shera, C.A., 2003. Mammalian spontaneous otoacoustic emissions are amplitude-stabilized Cochlear standing waves. *J. Acoust. Soc. Am.* 114, 244–262.
- Suga, N., Ma, X., Gao, E., Sakai, M., Chowdhury, S.A., 2003. Descending system and plasticity for auditory signal processing: neuroethological data for speech scientists. *Speech Commun.* 41, 189–200.
- Sumner, C.J., Wells, T.T., Bergevin, C., Sollini, J., Kreft, H.A., Palmer, A.R., Oxenham, A.J., Shera, C.A., 2018. Mammalian behaviour and physiology converge to confirm sharper cochlear tuning in humans. *Proc. Natl. Acad. Sci.* 115 (44), 11322–11326.
- Swaminathan, J., Krishnan, A., Gandour, J.T., 2008. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19 (11), 1163–1167.
- Talmadge, C.L., Long, G.R., Murphy, W.J., Tubis, A., 1993. New off-line method for detecting spontaneous otoacoustic emissions in human subjects. *Hear. Res.* 71, 170–182.
- Vihla, M., Kiviniemi, K., Salmelin, R., 2002. Auditory cortical activation in the Finnish and Swedish speaking Finns: a magnetoencephalographic study. *Neurosci. Lett.* 322, 141–144.
- Wan, I.K.K., Wong, L.L.N., 2002. Tympanometric norms for Chinese young adults. *Ear Hear.* 23 (5), 416–421.
- Whitehead, M.L., Kamal, N., Lonsbury-Martin, B.L., Martin, G.K., 1993. Spontaneous otoacoustic emissions in different racial groups. *Scand. Audiol.* 22, 3–10.
- Wong, P.C.M., Parsons, L.M., Martinez, M., Diehl, R.I., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J. Neurosci.* 24 (41), 9153–9160.
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Zatorre, R.J., 2003. Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nat. Rev. Neurosci.* 6 (7), 692–695.
- Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanism underlying melodic perception and memory for pitch. *J. Neurosci. Res.* 14 (4), 1908–1919.
- Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
- Zizz, C.A., Glattke, T.J., 1988. Reliability of spontaneous otoacoustic emission suppression tuning curve measures. *J. Speech Hear. Res.* 31, 616–619.