RESEARCH ARTICLE | DECEMBER 31 2015

Possible role of cochlear nonlinearity in the detection of mistuning of a harmonic component in a harmonic complex

Christophe Stoelinga; Inseok Heo; Glenis Long; Jungmee Lee; Robert Lutfi; An-Chieh Chang

AIP Conf. Proc. 1703, 090022 (2015)
https://doi.org/10.1063/1.4939420

Possible Role of Cochlear Nonlinearity in the Detection of Mistuning of a Harmonic Component in a Harmonic Complex

Christophe Stoelinga *, Inseok Heo*, Glenis Long†, Jungmee Lee*, Robert Lutfi * and An-Chieh Chang*

*Communication Sciences and Disorders, Univ. of Wisconsin-Madison, Madison, Wisconsin, USA
†Speech and Hearing Sciences, City University of New York, New York, New York, USA

Abstract. The human auditory system has a remarkable ability to "hear out" a wanted sound (target) in the background of unwanted sounds. One important property of sound which helps us hear-out the target is inharmonicity. When a single harmonic component of a harmonic complex is slightly mistuned, that component is heard to separate from the rest. At high harmonic numbers, where components are unresolved, the harmonic segregation effect is thought to result from detection of modulation of the time envelope (roughness cue) resulting from the mistuning. Neurophysiological research provides evidence that such envelope modulations are represented early in the auditory system, at the level of the auditory nerve. When the mistuned harmonic is a low harmonic, where components are resolved, the harmonic segregation is attributed to more centrally-located auditory processes, leading harmonic components to form a perceptual group heard separately from the mistuned component. Here we consider an alternative explanation that attributes the harmonic segregation to detection of modulation when both high and low harmonic numbers are mistuned. Specifically, we evaluate the possibility that distortion products in the cochlea generated by the mistuned component introduce detectable beating patterns for both high and low harmonic numbers. Distortion product otoacoustic emissions (DPOAEs) were measured using 3, 7, or 12-tone harmonic complexes with a fundamental frequency ($F0$) of 200 or 400 Hz. One of two harmonic components was mistuned at each $F0$: one when harmonics are expected to be resulted and the other from unresolved harmonics. Many non-harmonic DPOAEs are present whenever a harmonic component is mistuned. These non-harmonic DPOAEs are often separated by the amount of the mistuning (ΔF). This small frequency difference will generate a slow beating pattern at ΔF, because this beating is only present when a harmonic component is mistuned, it could provide a cue for behavioral detection of harmonic complex mistuning and may also be associated with the modulation of auditory nerve responses.

INTRODUCTION

In daily life, we are surrounded by sounds generated simultaneously by different sources. Sounds from different sources reach each ear as a single sound waveform but the human auditory system has a remarkable ability to hear them out as coming from separate sources. Many meaningful sounds, such as voiced-speech and music, are harmonic, inasmuch as their component frequencies are integer multiples of a common fundamental frequency ($F0$) and the patterns of harmonics play an important role source separation. The auditory system translates this harmonic structure to a pitch corresponding to $F0$. Hence, when two sounds with different $F0$ occur simultaneously the one will often be heard to segregate from the other.

Harmonic segregation has been studied extensively using a mistuning-perception paradigm wherein the listener is required to detect the mistuning of one component of a simultaneously-presented harmonic complex [2, 6, 7]. Two explanations of the harmonic segregation are given in these studies depending on whether the mistuned component is perceptually resolved (fall within a single processing bandwidth) or unresolved (harmonics falling in separate filters). Where the mistuned component is of a high harmonic number and unresolved, the segregation is thought to result from detection of modulation of the time envelope (roughness cue) resulting from the mistuning [4, 6]. But where the mistuned component is of a low harmonic number and resolved, the harmonic segregation is attributed to more centrally-located auditory processes, leading the harmonic components to form a perceptual group heard separately from the mistuned component. Here we consider an alternative account that attributes the harmonic segregation to detection of modulation at both high and low harmonic numbers.

Neurophysiological research provides evidence that such envelope modulations are available in the auditory nerve [11], cochlear nucleus [10] and inferior colliculus [12] of chinchillas, the temporal discharge pattern of a single unit neuron stimulated by mistuned components were modulated by interactions between adjacent components. Mul-

Mechanics of Hearing: Protein to Perception
AIP Conf. Proc. 1703, 090022-1–090022-5; doi: 10.1063/1.4939420

TABLE 1. Frequencies of the used stimuli

	Frequency range (Hz)			
$F0$	**mistuned**	**3 components**	**7 components**	**12 components**
200	400(=F2)	200-600	200-1400	200-2400
200	1800(=F9)	1600-2000	1200-2400	200-2400
400	1200(=F3)	800-1600	400-3000	400-4800
400	3200(=F8)	2800-3600	2000-4400	400-4800

tiunit activity recordings from Monkey's primary auditory cortex also reflect "beats" in the stimuli generated by the interactions between a mistuned component and adjacent components [1]. Specifically, we evaluate the possibility that distortion products generated in the cochlea by the mistuned component, and measured in the ear canal as Distortion Product OAEs (DPOAEs), introduce detectable beating, which may provide a cue for mistuning detection at both high and low harmonic numbers.

METHOD

Listeners

Five normal-hearing individuals without history of ear pathology were recruited for participation. All had pure-tone thresholds less than or equal to 20 dB hearing level (HL) at octave frequencies from $.25 - 8$ kHz. Normal middle ear function was verified clinically. The listeners provided written, informed consent and were compensated monetarily. All procedures were approved by the Institutional Review Board at University of Wisconsin - Madison.

Procedures

Equipment

Sound presentation and recording were done using the MATLAB programming environment, using some Psychophysics Toolbox extensions [3]. Sounds were presented via an RME fireface UCX audio interface and Rolls RA62c headphone amplifiers to Etymotic ER2 insert earphones coupled to an Etymotic ER10A three port microphone with the preamplifier set at +20dB. The amplified signal was input to the RME audio interface. Three separate ER2 sound units were connected to the ER10A to minimize equipment distortion. One ER2 presented the components with odd harmonic numbers, another one the components with the even harmonic numbers and the third one presented the mistuned component by itself. Thus, in the case where 3 components were used, each of these was presented via a separate ER2. This greatly reduces distortion components, such as the distortion products with frequency at $2f_1 - f_2$, generated when multiple components are from one transducer. Such system distortion was evaluated by making measurements in a CAP24 2 ml cavity. Some such distortion products (DPs) originating from one or more components, such as DP with frequency of $3f_1$, are seen in measurements made in the cavity and will also be present in the ear canal.

Stimuli

Distortion product otoacoustic emissions (DPOAEs) were obtained using 3, 7, or 12-tone harmonic complexes (each component at 65 dB SPL) with a fundamental frequency ($F0$) of 200 or 400 Hz. One of two harmonic components was mistuned for each $F0$: either a resolved harmonic (# 2 for $F0 = 200$ Hz and # 3 for $F0 = 400$ Hz) or an unresolved harmonics (# 9 for $F0 = 200$ Hz and # 8 for $F0 = 400$ Hz). Stimuli used are summarized in Table 1.

Recording

Each stimulus had a length of 4 seconds and was repeated 8 times. The phase of the components was randomized, but were constant for each stimulus. This was done in order to average the 8 recordings to reduce the noise floor. Note that DPOAEs will have stable phases and therefore add, whereas noise, possibly including spontaneous OAE's are not synchronous, and therefore consequently cancel upon averaging. These 4-s average sounds where then divided into 7 half-overlapping 1-s windows which were multiplied with a Hann window. The Fourier transform was taken of these signals and the resulting Power Spectra were averaged. The estimate of the Power Spectral Density (PSD) contains

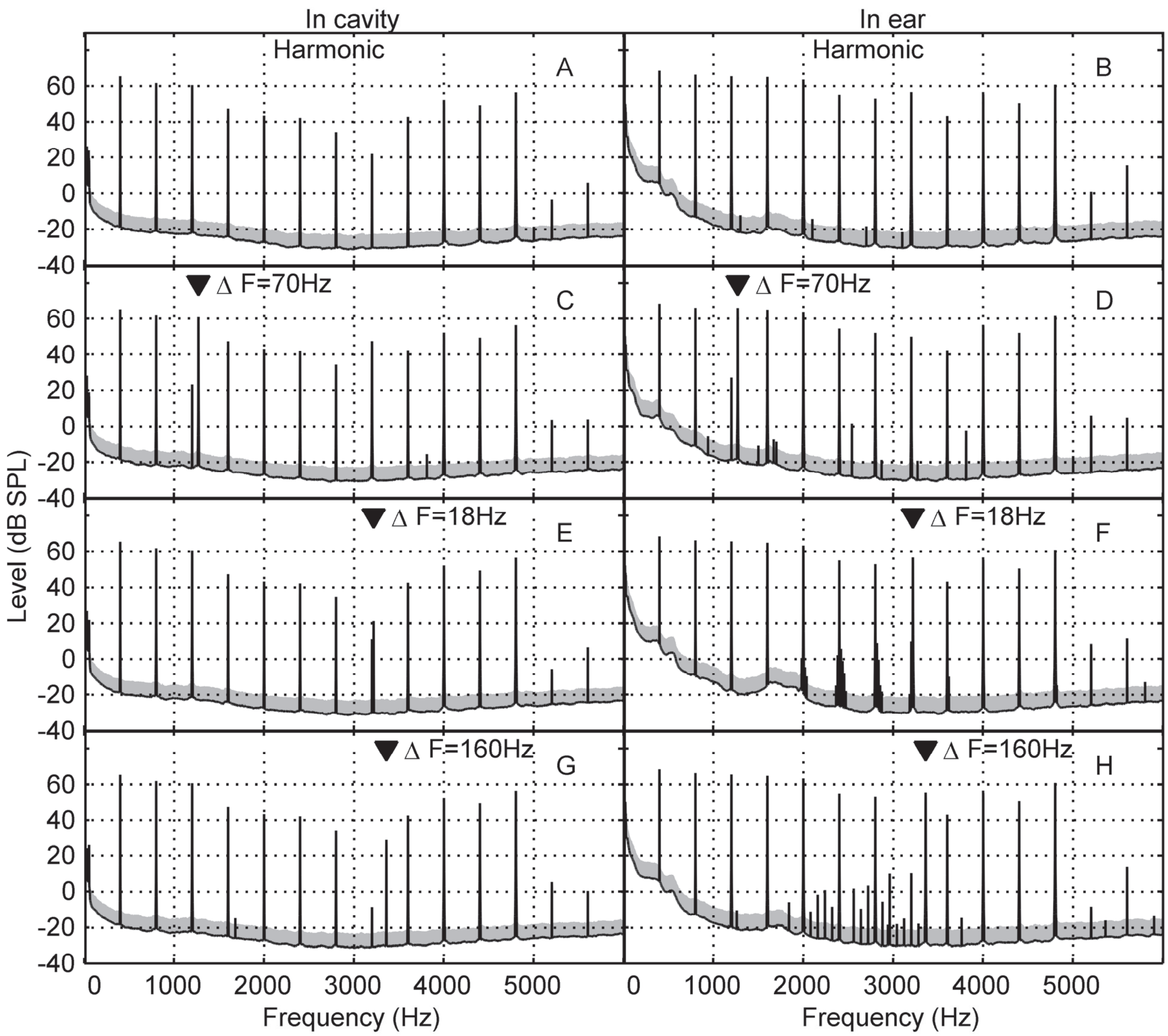

FIGURE 1. The spectra of recordings in cavity (left panels) and in the ear canal (right panels) of the 12-tone complex with $F0 = 400\,\text{Hz}$. The first row represents the measurements of an harmonic complex and the other rows represent the measurements of an harmonic complex where one component is mistuned by ΔF. The arrowhead in each panel indicates the mistuned component. The noise floor is indicated by the shaded region.

both noise and DPOAEs. The DPOAEs are relatively narrowband and sparse. They were distinguished from noise by estimating the envelope of the PSD using a 82-point median filter. This envelope was subtracted from the PSD resulting in Gaussian noise and a outliers more than 2 standard deviations away from the spectral envelope and fell at DPOAE frequencies were accepted as DPOAE.

RESULTS AND DISCUSSION

In the interest of space we are presenting a subset of the data in this paper. Other conditions will be included in the poster. The spectra of sound in a cavity (left panels) and in the ear canal (right panels) of a single typical listener are shown in Fig. 1 for 12-tone harmonic complexes. The fundamental frequency was 400 Hz. The measurements were made with the ER1OA placed near the entrance to the ear canal/cavity, so the levels of the harmonics appear unequal due to the resonances of the ear canal/cavity [9]. We compare the measurements in a cavity and in the ear canal to

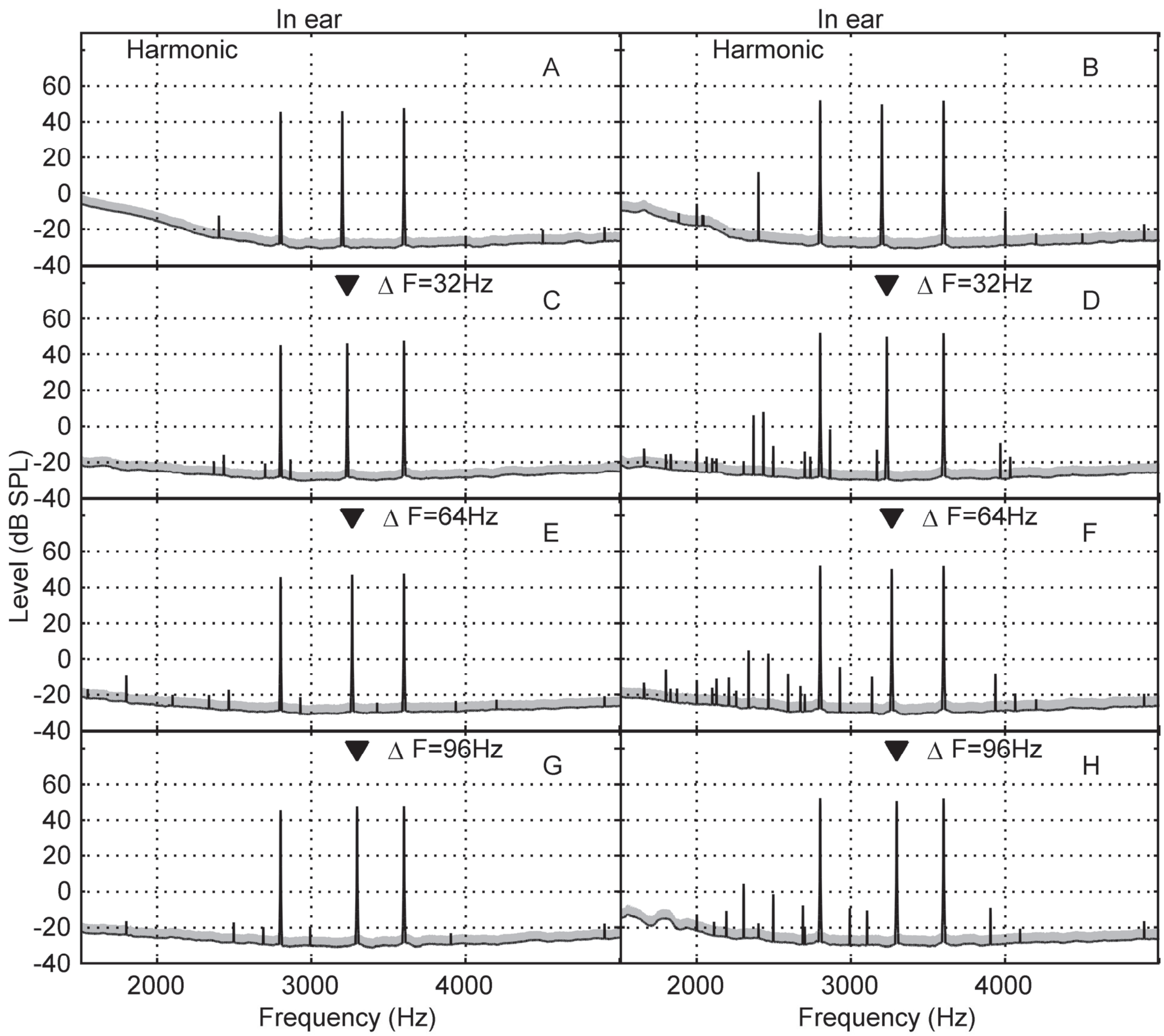

FIGURE 2. The spectra of recordings in the ears of two different listeners: one with higher level DPOAEs (right panels) and the other with low level DPOAEs (left panels). Three tone complexes were presented with frequencies f_1,f_2 and f_3 equal to 2800, 3200+ΔF and 3600Hz (the 7th, 8th and 9th harmonics of 400Hz). The two strongest DPOAEs seen in panels D, F and H have frequencies of $2f_1 - f_2$=2400-ΔF and $f_1 + f_2 - f_3$=2400+ΔF. When all components are harmonics, shown in panel A, the frequencies of both DP's, $2f_1 - f_2$ and $f_1 + f_2 - f_3$ fall at 2400 Hz. The arrowhead in each panel indicates the component which is mistuned. The noise floor is indicated by the shaded region.

determine which distortion products (DPs) detected in the ear canal are generated by the cochlea. The number of components used was based on psychoacoustic research using 12 components (c.f., [6, 7] but means that some system distortion was expected, but we were able to detect DPs in the ear canal which were not seen in the cavity. The panels in the first row (A and B) represent measurements obtained with a harmonic complex and the panels in the other rows (C-H) represent measurements obtained when one of harmonic component was mistuned by ΔF. The noise floor is indicated by the shaded region. The effect of mistuning of one harmonic component in the ear (panel D, F and H) leads to the generation of DPOAEs at frequencies determined by the amount of mistuning. These DPOAEs are absent in the corresponding measurements in a cavity (left panels) indicating that they reflect DPs generated in the cochlea. Several adjacent components have similar levels with a spacing dependent on the amount of mistuning which could produce a beating-like percept.

When three 3-tone harmonic complexes were used the DPs at low frequencies in ear were not detectable or reliable due to the high noise level due to body noise. In Fig. 2, 3-tone harmonic complexes are shown for different listeners: one with greater magnitude of DPOAEs (right panels) and the other with smaller magnitude of DPOAEs (left panels), DPOAEs generated when one component is mistuned depends on the amount of mistuning of the harmonic. For the listener with most prominent DPOAEs the two most prominent DPs, $2f_1 - f_2$=2400-ΔF and $f_1 + f_2 - f_3$=2400+ΔF, move in opposite directions when ΔF is increased. These two components generate a beating pattern whose frequency is predictable from ΔF. We chose to compare two subjects to reveal the intersubject variability in the levels of DPOAEs detected which may help explain some of the variability in psychoacoustic performance with similar stimuli from these subjects.

Because DPs generated in the cochlea must travel back through the middle ear before they are measured in the ear canal as DPOAE, the effective levels of the DPs in the cochlea are expected to be higher than the DPOAE in the ear canal. Indeed, psychophysical studies using DP cancellation and binaural masking level differences place the level of the most prominent DPs, $2f_1 - f_2$ and $2f_2 - f_1$, some 30 dB above the level measured in the ear canal [5]. This is confirmed by measures of forward and reverse travel of acoustic stimuli in human cadavers [8]. This would lead to a mere 20 dB difference between primaries and DPs in the cochlea at the frequency ratios used in this study.The families of generated DPs would potentially introduce detectable modulations on the basilar membrane when a harmonic component is mistuned, leading to psychoacoustic and physiological detection of harmonic mistuning.

ACKNOWLEDGMENTS

This research was supported by NIDCD grant R01 DC001280-21

REFERENCES

[1] Fishman YI, Steinschneider MS (2010) Neural correlates of auditory scene analysis based on inharmonicity in monkey primary auditory cortex. J Neurosci 30:12480–12494

[2] Hartmann WM, McAdams S, Smith BK (1990) Hearing a mistuned harmonic in an otherwise periodic complex tone. J Acoust Soc Am 88:1712–1724

[3] Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement 14

[4] Lee J, Green DM (1994) Detection of a mistuned component in a harmonic complex. J Acoust Soc Am 96:716–725

[5] Lutfi RA, Yost WA (1981) Measurement of the 2f1-f2 cubic difference tone with the binaural masking level difference. J Acoust Soc Am 69:216–222

[6] Moore BC, Peter RW, Glasberg BR (1985) Thresholds for the detection of inharmonicity in complex tones. J Acoust Soc Am 77:1861–1867

[7] Moore BC, Glasberg BR, Peter RW (1986) Thresholds for hearing mistuned partials as a separate tones in harmonic complexes. J Acoust Soc Am 80:479–483

[8] Puria S (2003) Measurements of human middle ear forward and reverse acoustics: implications for otoacoustic emissions. J Acoust Soc Am 113:2773–89

[9] Siegel, JH (2007) Calibrating otoacoustic emission probes. In Robinette MS, Theodore JG (eds) Otoacoustic Emissions: Clinical Applications, Stuttgart: Thieme, pp. 403–428

[10] Sinex DG (2008) Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. Hear Res 238:39–48

[11] Sinex DG, Guzik H, Li H, Sabes JH (2003) Responses of auditory nerve fibers to harmonic and mistuned complex tones. Hear Res 182:130–139

[12] Sinex DG, Sabes JH, Li H (2002) Responses of inferior colliculus neurons to harmonic and mistuned complex tones. Hear Res 168:150–162