

# 谐波复合音畸变产物产生模型

王 健<sup>1</sup>, 关 添<sup>2</sup>, 吴默村<sup>2</sup>, 杨文霞<sup>1</sup>

1. 深圳信息职业技术学院电子与通信学院, 广东 深圳 518172; 2. 清华大学深圳研究生院生物医学工程研究中心, 广东 深圳 518055

**【摘要】**由于人耳具有压缩非线性, 声音进入人耳后会产生畸变产物。为了深入地研究畸变产物对听觉音高感知的影响, 本文建立了一个畸变产物产生模型, 该模型产生的畸变产物频率低于谐波复合音主成分的频率。模型的实现过程包括: 计算谐波复合音的不同次幂、低通滤波、使用模拟退火遗传算法搜索最佳权重、加权求和。使用兴奋模式验证模型的正确性, 观察兴奋模式上的突出峰, 计算峰谷差, 匹配对应的畸变产物频率。结果发现畸变产物产生模型能够产生正确的畸变产物, 畸变产物能够增强谐波复合音的可分离性, 干扰不可分离谐波复合音的音高感知。本文提出在听觉音高感知实验中, 为了减小畸变产物对听觉音高感知的影响, 必须使用背景噪声掩蔽畸变产物。

**【关键词】**谐波复合音; 音高感知; 畸变产物; 模拟退火遗传算法

**【中图分类号】**R318.18

**【文献标识码】**A

**【文章编号】**1005-202X(2015)05-0639-04

## Model of generating distortion products of harmonic complex tone

WANG Jian<sup>1</sup>, GUAN Tian<sup>2</sup>, WU Mo-cun<sup>2</sup>, YANG Wen-xia<sup>1</sup>

1. School of Electronics and Communication, Shenzhen Institute of Information Technology, Shenzhen 518172, China;  
2. Research Center of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

**Abstract:** Because of the compression nonlinearity of human ears, distortion products (DPs) are generated after sounds stimulate the cochlear. A model of generating DPs was established in this paper to further study on the impacts of DPs on auditory pitch perception. The frequencies of DPs were lower than those of primary components of the harmonic complex tone. The implementations of the model included calculating different powers of the harmonic complex tone, low-pass filtering, searching the optimal weight by the simulated annealing genetic algorithm, weighting and summing the input signal. The excitation patterns were applied to verify the accuracy of this model. Salient peaks in excitation patterns were observed, and peak-to-valley differences were calculated, and corresponding frequencies of DPs were matched. Results showed that the model generated correct DPs which enhanced the resolvability of the harmonic complex tones, and disturbed the pitch perception of the unresolved harmonic complex tone. So DPs are needed to be masked by background noise to reduce the impact of DPs on auditory pitch perception in the auditory pitch perception simulation.

**Key words:** harmonic complex tone; pitch perception; distortion products; simulated annealing genetic algorithm

## 前言

谐波复合音由一系列正弦成分构成, 每个成分

的频率都是基频的整数倍。自然界的绝大部分声音都是谐波复合音, 例如: 语音、音乐、动物叫声等。由于人耳的压缩非线性, 当谐波复合音进入人耳后, 耳蜗会产生谐波复合音主成分之外的频率成分。这些成分称为畸变产物。

当频率为 $f_1$ 和 $f_2$  ( $f_1 < f_2$ ) 的两个正弦成分同时刺激耳蜗时, 产生的畸变产物频率为:  $f_2 + f_1$ ,  $f_2 - f_1$ ,  $2f_1 - f_2$ ,  $2f_2 - f_1$ ,  $3f_1 - 2f_2$  等。在这些畸变产物中, 人耳听不到频率大于 $f_1$ 的成分<sup>[1-4]</sup>, 但是能听到频率小于 $f_1$ 的成分, 即 $k(f_2 - f_1)$ 和 $f_1 - k(f_2 - f_1)$ , 其中 $k$ 为任意正整数<sup>[4-7]</sup>。此外, 随着 $k$

**【收稿日期】**2015-06-19

**【基金项目】**国家自然科学基金(31271056); 国家自然科学基金青年基金(81401539)

**【作者简介】**王 健(1985-), 女, 博士, 讲师, 主要研究方向: 听觉音高感知、语音信号处理。Tel: 0755-89226253; E-mail: wangj01@szit.com.cn。

**【通信作者】**关 添(1978-), 男, 博士, 清华大学副教授, 主要研究方向: 听觉机理、电子耳蜗、语音信号处理。Tel: 0755-26036771; E-mail: guantian@sz.tsinghua.edu.cn。

的增大,  $f_i - k(f_2 - f_1)$  的频率值减小, 声压水平也减小<sup>[4]</sup>。一般来说,  $k$  的最大值大约是 5 或 6<sup>[7]</sup>。研究表明: 主要的畸变产物频率为  $f_2 - f_1$ 、 $2f_1 - f_2$  和  $3f_1 - 2f_2$ <sup>[1-4, 6-7]</sup>。

在这些畸变产物中,  $2f_1 - f_2$  最为重要<sup>[4]</sup>。该畸变产物的声压水平几乎与谐波复合音主成分的声压水平无关, 而与主成分之间的频率间隔 (即  $|f_2 - f_1|$ ) 相关<sup>[3, 4, 8]</sup>。当主成分之间的频率间隔增大时,  $2f_1 - f_2$  的声压水平降低。研究表明: 当  $f_1$  和  $f_2$  分别为 1000 Hz 和 1100 Hz 时,  $2f_1 - f_2$  的声压水平大约比谐波复合音主成分的声压水平低 14 dB<sup>[4]</sup>。

对于畸变产物  $f_2 - f_1$ , 它的声压水平主要依赖于谐波复合音主成分的声压水平。只有当主成分的声压水平高于 50 dB SL 时, 畸变产物  $f_2 - f_1$  才能被人耳感知到<sup>[3-4]</sup>。如果谐波复合音的主成分只有  $f_1$  和  $f_2$ , 则畸变产物  $f_2 - f_1$  的声压水平比  $f_1$ 、 $f_2$  的声压水平大约低 20 dB~25 dB<sup>[4, 9]</sup>。

为了深入研究畸变产物对听觉音高感知的影响, 并用数学的方式表达听觉感知的非线性, 本文建立了一个畸变产物产生模型。

## 1 畸变产物产生机制

谐波复合音  $x(t)$  可用式 (1) 表示。其中,  $M$  代表谐波复合音包含的正弦成分个数,  $A_i$  和  $\Omega_i$  分别代表第  $i$  个正弦成分的幅度和角频率。

$$x(t) = \sum_{i=1}^M A_i \cos(\Omega_i t) \quad (1)$$

如式 (2) 所示, 在信号处理理论中, 若信号  $x(t)$  的成分两两相乘, 将会产生新的频率成分, 如原成分的倍频、和频、差频。该过程称为调制, 包括自调制和互调制。基于该理论, 本文提出: 畸变产物的产生可能源于谐波复合音各正弦成分的调制。

$$\begin{aligned} x^2(t) &= \left[ \sum_{i=1}^M A_i \cos(\Omega_i t) \right]^2 \\ &= \sum_{i=1}^M \frac{A_i^2}{2} [1 + \cos(2\Omega_i t)] + \\ &\quad \sum_{i=1}^{M-1} \sum_{j=i+1}^M A_i A_j [\cos((\Omega_i + \Omega_j)t) + \cos((\Omega_j - \Omega_i)t)] \end{aligned} \quad (2)$$

当谐波复合音  $x(t)$  刺激耳蜗时, 调制将产生  $x^2(t)$ 、 $x^3(t)$ 、 $x^4(t)$ 、 $x^5(t)$ ……。由式 (2) 可知, 若谐波复合音只包含两个正弦成分, 即  $M=2$  时, 则  $x^2(t)$  包含的畸变产物频率为  $2\Omega_1$ 、 $2\Omega_2$ 、 $\Omega_1 + \Omega_2$ 、 $\Omega_2 - \Omega_1$ 。

调制产物  $x^3(t)$  包含的畸变产物频率为:  $3\Omega_1$ 、 $3\Omega_2$ 、 $2\Omega_1 + \Omega_2$ 、 $2\Omega_2 + \Omega_1$ 、 $2\Omega_1 - \Omega_2$ 、 $2\Omega_2 - \Omega_1$ 。同理, 可以得到  $x^4(t)$ 、 $x^5(t)$ ... $x^N(t)$  以及相应的畸变产物。

统计  $x(t)$  的奇次幂和偶次幂, 可以发现:

(1) 对于奇次幂  $x^{2k+1}(t)$  ( $k \geq 1$ ), 频率小于  $f_1$  的畸变产物为  $2f_1 - f_2$ 、 $3f_1 - 2f_2$ ... $f_1 - k(f_2 - f_1)$ ; (2) 对于偶次幂  $x^{2k}(t)$  ( $k \geq 1$ ), 频率小于  $f_1$  的畸变产物为  $f_2 - f_1$ 、 $2(f_2 - f_1)$ ... $k(f_2 - f_1)$ 。

## 2 畸变产物产生模型

### 2.1 模型的建立

基于上述假说和数学表达, 本文建立了畸变产物产生模型 (图 1), 产生的畸变产物频率小于谐波复合音主成分的频率。

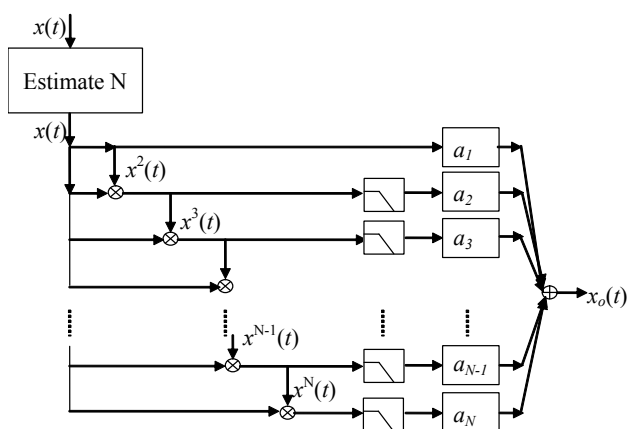


图 1 畸变产物产生模型

Fig.1 Model of generating distortion products

该模型的实现步骤为: (1) 估计谐波复合音  $x(t)$  的最大次幂  $N$ ; (2) 计算  $x(t)$  的  $n$  次幂  $x^n(t)$ ,  $n \in [2:N]$ ; (3) 将  $x^n(t)$  ( $n \in [2:N]$ ) 通过低通滤波器, 滤除频率大于  $f_1$  的高频畸变产物。由于频率小于  $f_1$  的畸变产物的最高频率为  $2f_1 - f_2$ , 因此低通滤波器的截止频率设置为声音主成分最小频率  $f_1$  和畸变产物最大频率  $2f_1 - f_2$  的平均值, 即  $(3f_1 - f_2)/2$ ; (4) 对  $x^n(t)$  按照式 (3) 进行加权求和, 权重为  $a^n$ 。

$$x_o(t) = \sum_{n=1}^N a^n x^n(t) \quad (3)$$

根据前人的研究结果,  $N$  的取值一般最大为 5<sup>[7]</sup>。

### 2.2 畸变产物声压水平的拟合

畸变产物产生模型的权重依赖于前人在心理学实验中测得的畸变产物声压水平。研究表明:  $k$  越大,  $k(f_2 - f_1)$  的频率值越大, 声压水平越小;  $k$  越大,  $f_1 - k(f_2 - f_1)$  的频率值越小, 声压水平越小<sup>[8]</sup>。因此, 畸变产物各成分的声压水平呈现一种两头高中间低的形状。因此, 本文假设: 频率小于  $f_1$  的畸变产物的声压

水平服从二次分布,并使用该假设和前人的心理声学实验数据<sup>[4, 9-10]</sup>拟合畸变产物的声压水平。

### 2.3 最佳权重 $\alpha^n$ 的确定

当 $x^n(t)$ 低通滤波后,畸变产物产生模型需要找到最佳权重并进行加权求和。寻找最佳权重的目的是使模型输出的畸变产物声压水平和心理声学实验测得的畸变产物声压水平相匹配。本文提出使用全局优化算法——模拟退火遗传算法来搜索最佳权重,其实现步骤是:(1)随机产生初始解,通过选择、交叉、变异等遗传操作产生一组新个体,找到最佳权重的大致范围;(2)在步骤(1)得到的大致范围内,对所产生的个体进行两层循环的模拟退火操作。外层循环控制温度的下降,内层循环在同一温度下进行状态的随机搜索。在每次搜索中,采用Metropolis准则(式(4))接收新解 $f(j)$ 。其中, $f(i)$ 是目标函数, $t_k$ 是温度。当新解 $f(j)$ 小于等于原解 $f(i)$ 时,新解代替原解;当新解 $f(j)$ 大于原解 $f(i)$ 时,新解以一定概率代替原解。

$$P = \begin{cases} 1 & f(j) \leq f(i) \\ \exp\left(-\frac{f(j)-f(i)}{t_k}\right) & f(j) > f(i) \end{cases} \quad (4)$$

(3)按照式(5)计算迭代误差。其中, $P_{m,i}$ 是模型产生的畸变产物的功率谱密度; $P_{r,i}$ 是拟合的对应畸变产物的功率谱密度; $K$ 是畸变产物的个数。如果迭代误差 $E$ 大于预设值,则重复步骤(1)和(2)。如果迭代误差 $E$ 小于等于预设值,则循环结束,以当前解作为最佳权重。

$$E = \sum_{i=1}^K (P_{m,i} - P_{r,i})^2 \quad (5)$$

### 2.4 模型验证

本文使用兴奋模式<sup>[11-12]</sup>来验证畸变产物产生模型的正确性。兴奋模式是声音在耳蜗基底膜的不同位置触发的兴奋分布。由于基底膜的每个位置对应一个听觉滤波器,因此,兴奋模式可以定义为基底膜上听觉滤波器组相对于中心频率的输出。兴奋模式可以用来表达谐波复合音各成分的可分离性。低次谐波( $\leq 6$ 次)是可分离的,在兴奋模式上具有突出的峰和谷,可分离谐波复合音可以利用较强的可分离性感知音高,音高感知结果较好。高次谐波( $> 8$ 次)是不可分离的,相应的兴奋模式曲线平滑,没有可见的峰和谷,不可分离谐波复合音只能利用微弱的时域包络信息感知音高,音高感知结果显著较差<sup>[13-14]</sup>。尽管如此,由于人耳的压缩非线性,不可分离谐波复合音在刺激耳蜗时,会再次引入畸变产物。由于人耳

可听的畸变产物具有较低的谐波次数,因此畸变产物可能会为音高感知提供较强的可分离信息。根据上述理论,本文将计算高次不可分离谐波复合音的兴奋模式来验证畸变产物产生模型的正确性。

兴奋模式的计算方法为<sup>[11-12]</sup>:使用Matlab中的PSD函数计算每个输入声音的功率谱密度,将功率谱密度输入到兴奋模式计算软件中。输入声音中的每个正弦成分将被一组听觉滤波器滤波,得到一条听觉滤波器组相对于对应中心频率的输出曲线。将所有正弦成分的输出曲线叠加在一起,则可以得到输入声音的兴奋模式。对兴奋模式的每个可见峰,用峰值减去相邻谷的平均值,得到峰谷差。如果峰谷差大于2 dB,则可见峰对应频率处的谐波是可分离的<sup>[15]</sup>,否则,可见峰对应频率处的谐波不可分离。

本文使用的输入声音刺激有4种:

(1)刺激a由频率分别为1000 Hz和1100 Hz的两个正弦成分构成,各成分的声压水平都是50 dB SPL。由于1000 Hz和1100 Hz的最大公约数是100,因此刺激a的基频为100 Hz,1000 Hz和1100 Hz分别对应基频的第10次和第11次谐波,属于不可分离谐波;(2)刺激b是将刺激a通过畸变产物产生模型得到的声音。该声音不但包含刺激a的成分,还将包含频率小于1000 Hz的畸变产物。这些畸变产物中包含可分离谐波和不可分离谐波;(3)刺激c是在刺激a中加入等阈值噪声(Threshold Equalizing Noise, TEN),该噪声在1000 Hz处的声压水平是40 dB/ERB<sub>N</sub>。TEN噪声在各个频率点的阈值相等<sup>[16]</sup>。ERB<sub>N</sub>是指在中等声压水平下,对听力正常的年轻人测得的听觉滤波器的等矩形带宽<sup>[17]</sup>。刺激c中包含1000 Hz和1100 Hz的不可分离谐波,但不包含畸变产物;(4)刺激d是在刺激b中加入40 dB/ERB<sub>N</sub>的TEN噪声。刺激d不但包含1000 Hz和1100 Hz的不可分离谐波,同时包含频率小于1000 Hz的畸变产物。

图2(a)~(d)分别为上述4个输入声音刺激的兴奋模式。刺激a包含1000 Hz和1100 Hz成分,但兴奋模式(图2a)在1000 Hz和1100 Hz处及附近都没有可见峰。这是因为:1000 Hz和1100 Hz是基频100 Hz的第10次和第11次不可分离谐波,因此在兴奋模式图中没有可见峰,与前人的结论一致<sup>[15]</sup>。

当刺激a通过畸变产物产生模型后,模型将在刺激a的基础上产生频率小于 $f_1$ 的畸变产物,得到刺激b。刺激b的兴奋模式如图2b所示:兴奋模式在100 Hz和200 Hz处有可见峰。这是因为100 Hz和200



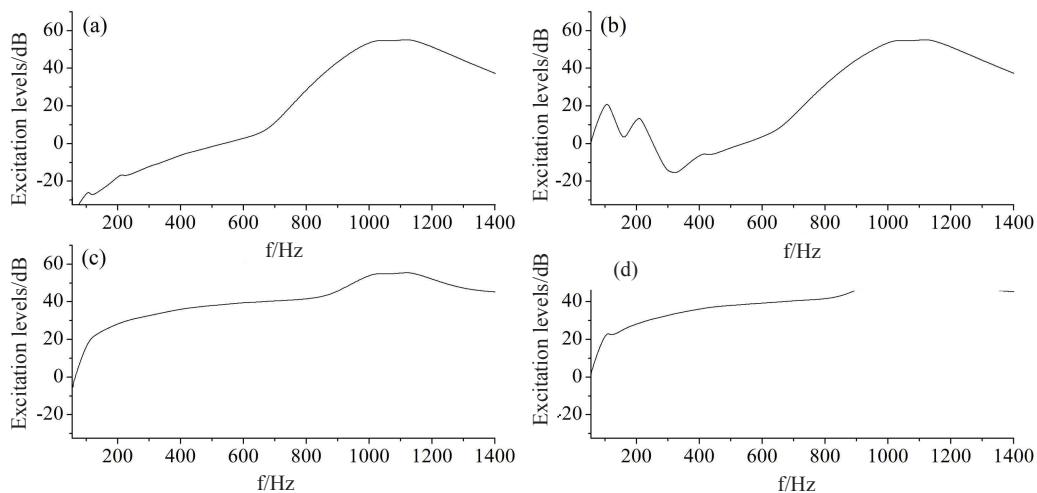


图2 不同声音刺激的兴奋模式

Fig.2 Excitation patterns of different stimuli

a: Stimulation a; b: Stimulation b; c: Stimulation c; d: Stimulation d

Hz处是畸变产物 $f_2-f_1$ 和 $2(f_2-f_1)$ ,二者的声压水平较大,并且是100 Hz的基波和二次谐波,具有很好的可分离性。因此,这两个畸变产物是可分离谐波,能够在兴奋模式上触发出突出的可见峰(峰谷差=9.8 dB)。由于这两个畸变产物的可分离性很强,因此能够增强原有不可分离谐波复合音的可分离性和音高感知。对于畸变产物 $2f_1-f_2$ 和 $3f_1-2f_2$ ,频率分别为900 Hz和800 Hz。尽管这两个成分的声压水平也很高,但是由于二者是基频100 Hz的第9次和第8次不可分离谐波,因此在兴奋模式上并没有对应的突出可见峰,只能诱发出微弱的音高,这与前人的结论一致<sup>[15]</sup>。

当不包含畸变产物的刺激a和包含畸变产物的刺激b置于40 dB/ERB<sub>N</sub>的TEN噪声中时,分别得到刺激c和刺激d。这两个声音的兴奋模式非常相似(图2c和图2d)。这是因为:40 dB/ERB<sub>N</sub>的TEN噪声掩蔽了畸变产物,减小了谐波的可分离性,降低了畸变产物对不可分离谐波复合音音高感知的影响。此外,刺激d和b都包含畸变产物,但刺激d在100 Hz和200 Hz处并没有显著峰,原因也是因为TEN噪声掩蔽了畸变产物。

上述结果与前人的结论相一致<sup>[4, 9, 10, 15]</sup>。因此,可以认为:本文建立的畸变产物产生模型可以产生正确的畸变产物,畸变产物将增强不可分离谐波复合音的可分离性,干扰音高感知结果。

### 3 结论

本文建立了一个畸变产物产生模型,使用兴奋

模式验证了模型的正确性,并探讨了畸变产物对不可分离谐波复合音音高感知的影响。主要内容为:(1)总结畸变产物与谐波复合音奇次幂、偶次幂的关系;(2)提出畸变产物产生机制可能源于信号的调制;(3)建立畸变产物产生模型;(4)使用模拟退火遗传算法搜索最佳权重;(5)使用兴奋模式研究畸变产物对不可分离谐波复合音音高感知的影响,对畸变产物产生模型进行验证。结果发现:畸变产物能够增强谐波复合音的可分离性。因此,本文提出:在谐波复合音的音高感知实验中,必须使用一定声压水平的背景噪声掩蔽畸变产物,以减小畸变产物对音高感知的影响。

### 【参考文献】

- [1] Kendall GS, Haworth C, Cádiz RF. Sound synthesis with auditory distortion products[J]. *Comput Music J*, 2014, 38(4): 5-23.
- [2] Haworth C. Composing with absent sound[C]. In *Proceedings of the international computer music conference*. 2011: 342-345.
- [3] Plomp R. Detectability thresholds for combination tones[J]. *J Acoust Soc Am*, 1965, 37: 1110-1123.
- [4] Goldstein JL. Auditory nonlinearity[J]. *J Acoust Soc Am*, 1967, 41: 676-689.
- [5] Hall JL, Schroeder MR. Monaural phase effects for two-tone signals [J]. *J Acoust Soc Am*, 1972, 51: 1882-1884.
- [6] Smoorenburg GF. Combination tones and their origin[J]. *J Acoust Soc Am*, 1972, 52: 615-632.
- [7] Smoorenburg GF. Audibility region of combination tones[J]. *J Acoust Soc Am*, 1972, 52: 603-614.
- [8] Robles L, Ruggero MA, Rich NC. Two-tone distortion on the basilar membrane of the chinchilla cochlea[J]. *J Neurophysiol*, 1997, 77: 2385-2399.

(下转 663 页)