doi: 10.7690/bgzdh.2014.04.009

基于谱减法及频谱方差的语音端点检测方法

刘新飞，周辉
（装备学院信息装备系，北京 101416）

摘要：为提高在强背景噪声条件下语音端点检测的性能，提出一种将改进的谱减法和频谱方差相结合的方法。先利用谱减法对带噪语音信号进行增强处理，然后使用基于频谱方差的方法进行检测，并以一男一女对 10 个数字和 26 个英文字母各发音 2 次的录音作为样本进行仿真验证。实验结果表明：该方法计算简单，可靠性高，能有效减少背景噪声，提高语音信号的信噪比，在较低的信噪比下仍能比较准确地检测到语言信号的端点。

关键词：谱减法；频谱方差；端点检测
中图分类号：TP216　　**文献标志码：**A

A Voice Activity Detection Algorithm Based on Spectral Subtraction and Spectrum Variance

Liu Xinfei, Zhou Hui
(*Department of Information Equipment*, *College of Equipment*, *Beijing* 101416, *China*)

Abstract: For improving performance of voice activity detection (VAD) under big background noise, proposes an improved method by combining modified spectral subtraction and spectral variance. The proposed scheme employs modified spectral subtraction to enhance the speech firstly, and then detect based on spectral variance. And at the same time, a large number of simulation experiments have been done to confirm the validity of the scheme which take the pronunciations of 10 digits and 26 letters recorded each one twice by a man and an woman as testing samples. Experiment result shows that the method simple calculation and high reliability, it can reduce background noise effectively, increase signal-to-noise ratio of voice signal, can accurately detection voice signal endpoint under low signal-to-noise ratio.

Keywords: spectral subtraction method; spectrum variance; endpoint detection

0　引言

在语音信号处理中，语音活动检测(voice activity detection，VAD)是一个非常重要的环节。在孤立词自动语音识别系统中，语音端点检测的不准确是识别错误最主要的来源。在现实环境中，存在各种各样的突发噪声，使得背景环境总是非平稳变化[1]。由于语音信号具有短时平稳特性，在传统的端点检测方法中，主要是使用短时能量和短时过零率等参数把语音信号和背景噪声加以区分。在信噪比较高时，这些方法都能准确检测到语音的起止点。但是，在非平稳噪声环境下，这些需要依靠经验来设置判决阈值或噪声门限方法的弊端就显现出来了[2]。低信噪比条件下，背景噪声对这些短时特性参数的影响非常大，不能完全得到纯净语音的特征，算法的抗噪声性能很差，导致整个识别系统性能下降。目前还有一些改进的语音端点检测技术，准确性好、稳健性较强，虽然检测精度得到了一定的提高，但它们往往计算量大、复杂度高，不利于语音信号的实时处理[3-4]。

在低信噪比条件下，传统的端点检测方法侧重于提取更稳健的语音特征参数，而对噪声环境的影响考虑的较少；因此，端点检测的效果并不理想。谱减法是一种有效的语音增强技术，是先对噪声的频谱进行估计，而后通过“谱相减”去除噪声段的技术。其计算复杂度低，实时性强。“频谱方差”是某一帧信号的各频带能量间的方差。实际环境中，语音信号各频带间变化比较剧烈，而背景噪声频谱中，各频带之间变化很平缓，可利用这一特性来区分语音部分和噪声部分[5]；因此，笔者先利用谱减法对带噪语音信号进行增强处理，再使用基于频谱方差的方法进行端点检测，可以有效提高检测效率。

1　谱减法语音增强

1.1　理论分析

谱减法因其算法简单，去噪效果好，已成为最流行的语音降噪技术之一。基本谱减法的思想是假设语音信号和噪声信号在时域和频域上都是加性的，且互不相关。语音信号的功率谱在整个频域内是不断变化的，而噪声功率谱在整个频域内是平稳变化的，可以看作常量。先估计出带噪语音信号的噪声功率谱，然后从带噪语音功率谱中减去噪声功

收稿日期：2013-11-11；修回日期：2013-12-26
作者简介：刘新飞(1991—)，男，甘肃人，硕士，从事信息采集与处理研究。

率谱估计值，则可以获得降噪后较纯净的语音信号，从而达到语音增强的目的。由于噪声可以分为加性噪声和非加性噪声，而非加性噪声可通过同态变换变为加性噪声；因此，笔者主要对加性噪声中最普遍的宽带噪声进行降噪分析与处理。图 1 为谱减法去噪的基本原理框图[6]。

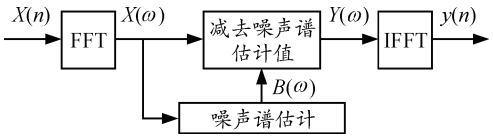

图 1　基本谱减法原理框图

图 1 中，假设带噪信号为

$$x(n)=s(n)+b(n) \tag{1}$$

其中：$s(n)$为纯净语音；$b(n)$为背景噪声，假设背景噪声是平稳加性且与语音信号是不相关的随机过程。尽管语音信号是时变信号，但是它具有短时平稳性，在很短的时间间隔内保持平稳不变。根据这一特性，可以利用短时傅里叶变换(short time fourier transform，STFT)对语音信号进行分析[7]。在时域对语音信号进行加窗分帧，如下式：

$$x_{\mathrm{pl}}(n)=w[pL-n](s(n)+b(n)) \tag{2}$$

其中 L 是窗的长度，而 p 是一个整数。对式 (2) 进行傅里叶变换可得：

$$X(pL,w)=S(pL,w)+B(pL,w) \tag{3}$$

式中 $X(pL,w)$、$S(pL,w)$和 $B(pL,w)$分别是 $x(n)$、$s(n)$和 $b(n)$在窗长为 L 时计算得到的短时傅里叶变换。因此，$x(n)$傅里叶变换的平方可以表示为：

$$|X(pL,w)|^2=|S(pL,w)|^2+|B(pL,w)|^2+ S^*(pL,w)B(pL,w)+S(pL,w)B^*(pL,w) \tag{4}$$

由于语音信号和加性噪声相互独立的，因此有：

$$|X(pL,w)|^2=|S(pL,w)|^2+|B(pL,w)|^2 \tag{5}$$

显而易见，如果能够得到$|S(pL,w)|^2$的估计值，同时忽略语音和噪声的相位差别，对分帧后的语音信号就有 $\hat{S}_i(pL,w)=|S_i(pL,w)|\mathrm{e}^{j\angle X_i(pL,w)}$，$|S_i(pL,w)|$表示对第 i 帧信号幅度的估计值。这样，就可以对原始语音信号进行恢复和重构[6]。

由于平稳噪声的功率谱在“寂静段”和“语音段”期间可以认为基本没有发生变化，故可通过“寂静段”(这一段里没有语音只有噪声)来估计第 i 帧信号中噪声的功率谱 $\hat{S}_{b(i)}(\omega)$，从而有：

$$|\hat{S}_i(pL,w)|^2=\begin{cases}|X_i(pL,w)|^2-\hat{S}_{b(i)}(w) & \text{if } |X_{(i)}(pL,w)|^2-\hat{S}_{b(i)}(w)\geqslant 0\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

其中$|X_i(pL,w)|^2$表示第 i 帧带噪语音信号的功率谱。这样相减之后得到的功率谱即可认为是较为纯净的语音功率谱，从这个功率谱可以计算出语音信号的幅度，再结合之前保留的相位可以得到降噪后的语音时域信号 $\hat{s}(n)$，则 $\hat{s}(n)=\mathrm{IFFT}[|\hat{S}(pL,w)|\ \mathrm{e}^{j\angle X(pL,w)}]$。

1.2　改进的谱减法

传统的谱减法忽略了语音信号相位的变化，从带噪语音信号功率谱中减去噪声信号功率谱估计值，从而达到语音增强的目的。在基本的谱减法中，假定背景噪声是平稳变化的，即带噪语音中的噪声具有和语音段开始前的“寂静段”相同的统计特性，且在整个语音段中保持恒定值不变；因此，在谱相减过程中，使用“寂静段”噪声的统计平均代替当前帧的噪声。

现实中，这种理想的平稳噪声几乎是不存在的，背景噪声估计的准确性直接影响着语音增强的效果。大多数情况下，在噪声段，后一帧的背景噪声可以直接用前一帧的信号来估计。但在语音段，由于噪声和语音信号叠加在一起，如果仅使用“寂静段”噪声的统计平均作为语音段背景噪声的估计值，在进行谱相减时，就会残留一些剩余的谱峰。这样，经过谱减增强后，语音信号就会产生一些新的残余噪声。在频域上就成为一些离散的谱峰，相应地在时域上就呈现出一些类正弦信号的叠加，这样形成的残余噪声就是“音乐噪声”。“音乐噪声”的产生，明显地降低了语音的可懂度，因此有必要不停地对噪声的估计值进行更新，从而提高去噪的效果。

由于语音信号具有短时平稳性，即在一段很短的时间间隔内，语音信号保持相对稳定的特征。因而可以对语音信号进行分帧处理，一般帧长取 10～30 ms。对于平稳加性噪声，在谱减过程中，可以通过不断更新噪声估计值来提高去噪效果。因此，可以对短时噪声谱估计方法进行改进，数学表达式为：

$$\hat{S}_{b(i)}(w)=\lambda\hat{S}_b(w)+(1-\lambda)\hat{S}_{b(i-1)} \tag{7}$$

其中：$\hat{S}_{b(i)}(w)$表示第 i 帧噪声信号的功率谱估计值；$\hat{S}_{b(i-1)}(w)$为第 $i-1$ 帧噪声信号的功率谱估计值；$\hat{S}_b(w)$表示语音刚开始时“寂静段”的平均功率谱；λ为平滑系数，由噪声的特性来决定，反映了噪声频谱的变化趋势，取值一般为[0.1,0.9]。高斯白噪声属于平稳随机过程，随机变化的幅度比较小，可以取 λ=0.1。而 Babble 噪声是由多说话人形成的嘈杂人群噪声，噪声随机变化的幅度比较大，经常会出

现某点的噪声分量非常大，此时应该“多减”，根据噪声本身的特性，可以取 λ=0.5。

用笔者所述方法对实际语音信号进行仿真验证。实验中，利用话筒在安静环境下录得一段语音，做为仿真实验中所需的纯净语音信号。经 8 kHz 采样，PCM 编码，量化为数字信号，单声道。噪声为高斯白噪声和 Babble 噪声。在 Matlab 中混合生成 3 种不同信噪比（−5, 0, 5 dB）的带噪语音，用 Matlab7.0 仿真实现。图 2 为在高斯白噪声条件下，加噪后带噪语音信噪比为−5 dB 时，改进的谱减法与基本谱减法去噪效果对比。从图 2 中可以看出，通过对谱减过程中的噪声估计值进行实时更新，去噪效果有了很大提高。

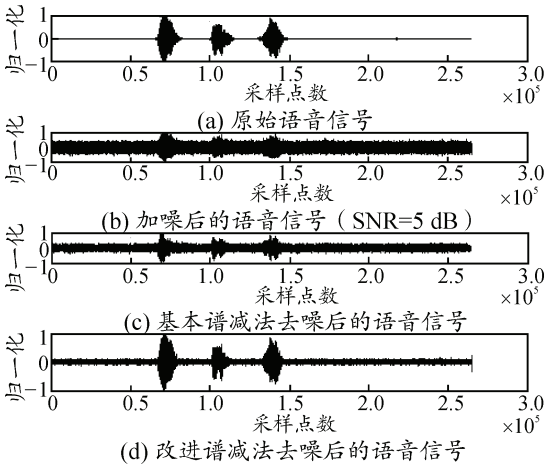

图 2 白噪声条件下谱减去噪效果

表 1 给出了在语音信噪比分别为 −5, 0, 5 dB 的情况下，在高斯白噪声和 Babble 噪声条件下，基本谱减法增强效果与文中提出的方法得出的实验结果相比较。实验表明：笔者提出的语音去噪方法使噪声得到很明显的抑制，提高了语音质量。

表 1 不同噪声环境下语音去噪实验结果

输入信噪比 (SNRin)	SNRout（高斯白噪声条件下）	SNRout（Babble 噪声条件下）
−5	7.039 5	1.327 6
0	9.881 1	3.960 4
5	13.731 5	7.065 2

2 端点检测算法

2.1 频谱方差

语音信号和噪声信号的频谱特性相差很大。对于确定信号，可以通过对其自相关函数进行傅里叶变换而得到功率谱。一个随机序列 $s(n)$，其频谱谱线越集中说明其自相关性越强。语音信号各个频段的幅度值变化比较剧烈，并且其有效部分集中在 300～3 400 Hz 这一频段。在短时间内，语音信号往往具有很强的自相关性，甚至具有周期性。这样的信号，其频谱会集中在一个很窄的频段内，而其他频段的功率谱值很小。

实际语音环境中，广泛存在白噪声，几乎分布于音频处理系统的所有频段。白噪声功率谱密度为常数，也就是说白噪声在各个频段上的功率是一样的，并且相关性极低。

因此，完全可以有效地利用语音和噪声频谱分布特性的这些差异，有效区分语音信号和背景噪声信号。如果能够用一个具体的变量来表示这些频谱特性的差异，就可以通过对这个量设定阈值来判定语音信号的起止点。如前所述，如果计算语音信号的频谱方差，会得到一个较大的值。而白噪声分布的频带宽，功率谱则较为平坦，且谱值一般都比较小，计算其频谱方差，得到的值明显小于语音信号数值。因此，选用适当的方法分析语音和噪声的频谱方差，就能判定语音信号的起止点。

假设输入语音信号为 $x(n)$，加窗分帧处理后得到的第 i 帧信号为 $x_i(m)$，每帧的长度为 N，则 $1 \leqslant m \leqslant N$。首先要对 $x_i(m)$ 进行离散傅里叶变换（discrete fourier transformation，DFT）：

$$X_i(k) = DFT[x_i(m)] = \sum_{m=0}^{N-1} x_i(m) W_N^{mk} (k = 0,1,\cdots,N-1) \tag{8}$$

由于 DFT 算法计算量太大，实际应用中通常使用其快速算法（fast fourier transform，FFT）。

用矢量 X 来表示频域分量的组合：

$$X = \{x_i(1), x_i(2), \cdots, x_i(N)\} \tag{9}$$

上式中各个分量的均值为：

$$E = \frac{1}{N} \sum_{j=0}^{N-1} x_i(j) \tag{10}$$

则频谱方差定义为：

$$D = \frac{1}{N} \sum_{j=0}^{N-1} [x_i(j) - E]^2 \tag{11}$$

从式 (11) 中可以看出，频谱方差反映的是语音信号频域能量变化的波动程度。D 值越大，信号能量越大，频率变化越剧烈。语音信号在频域内集中在 300～3 400 Hz 这一频段，在其他频段的分量很小，因此频谱方差会比较大。而对于白噪声信号，它的能量较小，功率谱分布于所有频段，起伏也比较平缓，因此频谱方差较小。由此，就可以根据频谱方差来判定语音端点[5]。

使用文献[5]方法进行端点检测时，需要确定一个合适的门限值 D_n。语音信号开始时“寂静”部分一般是噪声信号，通常的办法是将开始阶段的 10～20 帧信号作为噪声模型，计算它们的频谱方差，求均值，判决门限 D_n 可设置为这个均值的 3～5 倍。根据这个门限，即可将噪声和语音信号区分开。

2.2　算法原理与步骤

传统的语音端点检测方法通过计算语音信号的短时特征参数，可以有效地检测出语音的起止点。语音的时域特征参数主要有短时能量和平均幅度、短时过零率、短时自相关函数等，频域特征参数有语音的谱熵、倒谱、频谱方差等[7]；但是，在低信噪比环境中，这些方法的检测准确率很低，甚至失效；因此，笔者提出基于谱减法和频谱方差的端点检测方法。把谱减增强算法与频谱方差结合起来，使用改进的谱减法对带噪语音信号进行降噪处理，有效地提高了语音信号的信噪比，之后通过计算增强后语音信号的频谱方差，可以更加准确地检测出语音信号的端点。

具体算法的实现步骤为：

1) 对输入语音信号 $x(n)$ 进行预处理，包括预加重、加窗和分帧。

2) 利用短时傅里叶变换对带噪语音信号进行分析，计算得到每一帧信号的功率谱 $|X_i(pL,w)|^2$。

3) 用 1.2 节方法进行噪声谱估计，得到每一帧信号的噪声谱估计值 $\hat{S}_{b(i)}(\omega)$，并用 $|X_i(pL,w)|^2$ 减去 $\hat{S}_{b(i)}(\omega)$，从而得到各帧较为纯净的语音功率谱 $|\hat{S}_i(pL,w)|^2$。

4) 首先对 $|\hat{S}_i(pL,w)|^2$ 进行开方运算，可得到谱相减后每帧信号幅度的估计值，再结合相位信息，可得到谱减增强后各帧语音信号 $\hat{S}_i(pL,w)$。然后通过叠加、滤波并且平滑处理后进行傅里叶反变换，就可以重构出降噪后语音信号 $\hat{s}(n)$。

5) 用 2.1 节方法计算 4)中增强后各帧语音信号 $\hat{S}_i(pL,w)$ 的频谱方差 D_i，并选取增强后语音信号前 10 帧，计算其频谱方差，取其 5 倍作为初始判决门限 D_n 进行端点检测。在检测过程中，当计算所得的频谱方差值连续 5 帧单调增加时，作为语音部分的起点；否则，认为是背景噪声段。之后，对这 5 帧信号的频谱方差值进行加权平均，取其均值作为新的判决门限继续进行语音的端点检测。进入语音部分后，停止对判决门限的更新，以免出现误判。

6) 输出端点检测结果。

语音端点检测方法的主要流程如图 3 所示。

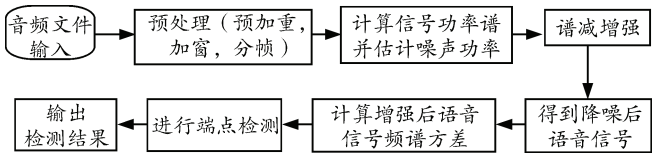

图 3　端点检测方法流程

3　仿真分析

笔者在计算机上对提出的方法进行了仿真验证，语音信号由一男一女对 10 个数字和 26 个英文字母各发音 2 次的录音组成的语音数据库样本提供。端点检测过程基于 Matlab7.0 实现。噪声使用 NOISEX-92 数据库中的 Babble 嘈杂人群噪声信号以及由 Matlab 生成的高斯白噪声，并将语音和噪声按比例线性相加生成不同信噪比(−5, 0, 5, 10, 15 dB)的带噪语音。

为了验证文中提出的基于谱减法和频谱方差的语音端点检测方法的有效性，将其与目前使用最广泛的，由 L.R.Rabiner 提出的基于短时能量和过零率的双门限检测方法[8]和文献[5]中的基于频谱方差的检测方法同时进行比较。图 4 和图 5 是在不同噪声条件下 3 种端点检测方法的检测结果。准确率的计算式为

$$\text{准确率}=\frac{\text{检测的正确的语音样本数}}{\text{总的语音样本数}}\times 100\% \qquad (12)$$

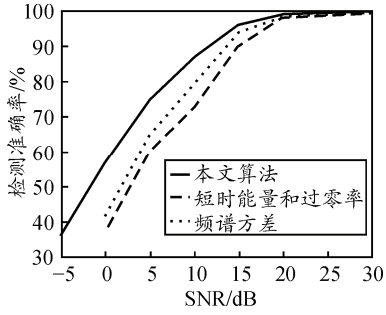

图 4　高斯白噪声条件下 3 种方法的检测结果对比

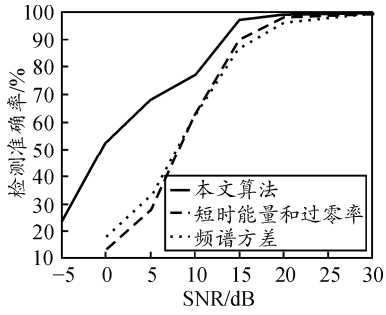

图 5　Babble 噪声条件下 3 种方法的检测结果对比

从图 4 和图 5 可以看出，经过谱减增强后的语音信号，进行端点检测时准确率明显高于基于短时能量和过零率方法以及基于频谱方差的检测方法。

在信噪比较低时，基于短时能量和过零率的检测方法及利用频谱方差的方法检测正确率很低，甚至在信噪比为−5 dB 时已经失效。而笔者提出的通过先利用谱减去噪再使用频谱方差进行检测的方法优势更加明显。随着信噪比的提高，3 种方法的检测正确率都有较大的提高，并趋于一致。在白噪声条件下，基于频谱方差的方法明显优于基本的短时能量和过零率方法。而在 Babble 噪声条件下，当信噪比较低时，基于频谱方差的方法尚有一定的优势，但是随着信噪比的提高，此方法检测正确率甚至低于基本的短时能量和过零率方法。这也符合嘈杂人群噪声的特性。

分析图中结果可见，在低信噪比时，笔者提出的基于谱减法和频谱方差的端点检测方法仍能保持较高的检测准确率。

4 结束语

针对传统语音端点检测方法存在的缺陷，笔者提出了基于谱减法和频谱方差的端点检测方法。通过改进的谱减增强算法，提高了语音信号的信噪比，再结合基于频谱方差的端点检测方法，能够大大提高语音端点检测的性能。通过理论分析与实验研究，证明了该算法快速有效，在低信噪比情况下也有良好的性能，具有一定的鲁棒性。

参考文献：

[1] Savoji M. H. A robust algorithm for accurate endpointing of speech[J]. Speech Commun, 1989, 8: 45−60.

[2] 吕卫强, 黄荔. 基于短时能量加过零率的实时语音端点检测方法[J]. 兵工自动化, 2009, 28(9): 69−70.

[3] Wu GinDer, Lin ChinTeng. A Recurrent Neural Fuzzy Network for Word Boundary Detection in Variable Noise-Level[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cybernetics, 2001, 31(1): 84−97.

[4] Junqua J. C, Mak B., Reaves B. A robust algorithm for word boundary detection in the presence of noise[J]. IEEE Trans, Speech Audio Processing, 1994, 7(2): 406−412.

[5] 刘玉珍, 连自锋. 基于频谱方差的抗噪声语音端点检测算法[J]. 计算机仿真, 2010, 27(9): 337−340.

[6] Chen Zhixin. Simulation of Spectral Subtraction Based Noise Reduction Method[J]. International Journal of Advanced Computer Science and Applications, 2011, 2(8): 30−32.

[7] 张雪英. 数字语音处理及 Matlab 实现[M]. 北京: 电子工业出版社, 2010: 1−7.

[8] Lawrence Rabiner, Sambur M. R. An Algorithm for Determining the Endpoints of Isolated Utterances[J]. The Bell System Technical Journal, 1975, 54(2): 297−315.

****

（上接第 25 页）

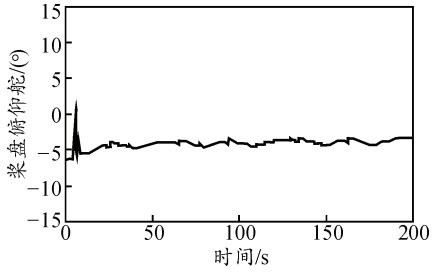

图 11 爬升段无人机桨盘俯仰舵(升降)变化曲线

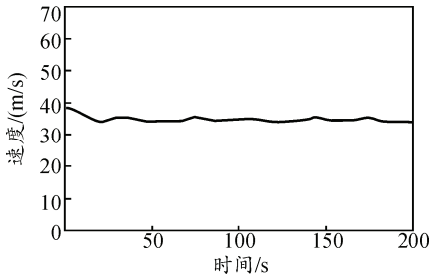

图 12 爬升段无人机速度变化曲线

仿真结果显示：爬升阶段自转旋翼无人机俯仰角基本跟踪指令信号，姿态逐渐稳定，爬升率为 5 m/s，爬升状态符合设计要求。

6 结论

笔者设计了一种自转旋翼无人机半实物仿真系统，能够真实地模拟自转旋翼无人机飞行控制系统运行环境，具有较高的仿真置信度和可靠度，对于加快研制进度、评估飞行品质、节省研制费用有极其重要的意义。另外，通过修改无人机六自由度模型、仿真模型、控制律及参数等，可将该系统应用在其他无人机研制过程中[10]。

参考文献：

[1] 刘亮亮, 等. 无人机半实物仿真系统研究[J]. 兵工自动化, 2008, 27(3): 44−45.

[2] 王行仁, 等. 飞行实时仿真系统及技术[M]. 北京: 北京航空航天大学出版社, 1998: 12−25.

[3] 沈永璋, 等. 飞行控制系统数字仿真[J]. 南京航空航天大学学报, 1985, 12(1): 22−26.

[4] 李柯, 等. 某型地空导弹半实物仿真系统训练效果的评估[J]. 兵工自动化, 2008, 28(12): 85−88.

[5] 熊光楷, 等. 先进仿真技术与仿真环境[M]. 北京: 国防工业出版社, 1997: 12−14.

[6] 魏凯, 李志国, 马存旺. 自转旋翼无人机技术与发展前景[J]. 飞航导弹, 2012, 12(12): 37−40.

[7] 朱清华, 李建波, 倪先平, 等. 可跳飞自转旋翼飞行器推/升力系统参数优化[J]. 航空动力学报, 2008, 23(1): 75−80.

[8] 陈睿璟, 刘晓飞, 罗珊. 无人直升机自主飞行的全数字仿真[J]. 兵工自动化, 2012, 31(10): 33−34.

[9] 李英杰, 贾燕军, 李相民. 近程放空导弹拦截巡航导弹的建模与仿真[J]. 兵工自动化, 2010, 29(12): 38−41.

[10] 闫宇壮, 等. RTX 在半实物仿真中的软件开发方法[J]. 兵工自动化, 2006, 25(9): 89−90.