

线谱频率及差分线谱频率参数相关性分析

韩笑蕾¹, 赵晓群¹, 方腾龙¹, 贾晓光²

(1. 同济大学 电子与信息工程学院, 上海 201804; 2. 燕山大学 学生工作处, 河北 秦皇岛 066004)

(hxl_regina@hotmail.com)

摘要:根据英语/汉语男女声线谱频率(LSF)参数及差分 LSF 参数帧内相关性统计结果,提出适合于 LSF 参数及差分 LSF 参数的分裂矢量量化(SVQ)分组方案。实验表明,在不考虑码书大小的情况下使用 SVQ 量化 10 阶 LSF 参数时,(4,6)分组的量化效果较优,否则(4,2,4)或(4,4,2)分组的量化效果较优。通过相关程度分布表清晰表明,至少 68% 的差分 LSF 参数在帧内呈微相关,有效减少了 LSF 参数的帧内冗余信息。随后采用 DSQ 和多种分组的 EEDSVQ 对差分 LSF 参数进行量化,结果表明差分 LSF 的量化性能优于 LSF 参数的量化性能。在语音编码中,采用差分 LSF 参数代替 LSF 参数作为模型参数,可在保持相同语音质量的情况下进一步降低编码速率。

关键词:线谱频率;差分 LSF 参数;帧内相关性;帧间相关性;分裂矢量量化

中图分类号: TN912.3 **文献标志码:** A

Correlation analysis of LSF and differential LSF parameters

HAN Xiao-lei¹, ZHAO Xiao-qun¹, FANG Teng-long¹, JIA Xiao-guang²

(1. School of Electronics and Information, Tongji University, Shanghai 201804, China;

2. Student Affairs Office, Yanshan University, Qinhuangdao Hebei 066004, China)

Abstract: In the light of intra-frame correlation of Line Spectrum Frequencies (LSFs) and Differential LSFs (DLSFs) from English/Chinese female/male speech database, an optimal partition scheme for LSFs and DLSFs was proposed. The experimental results show that if the size of codebook is not limited, dividing 10-order LSF vector into two sub-vectors of (4, 6) can get better quantization performance, otherwise dividing LSF vector as (4, 2, 4) or (4, 4, 2) can get better quantization performance. The intra-frame correlation between DLSFs is significantly smaller than that between LSFs, and at least 68% of DLSFs has feeble intra-frame correlation. DLSFs were quantized by DSQ and EEDSVQ, and the experiments show that the quantization performance of DLSFs is better than that of LSFs. In speech coding systems, adopting DLSFs instead of LSFs can get less spectral distortion and reach high quality speech at lower bitrates.

Key words: Line Spectrum Frequency (LSF); Differential Line Spectrum Frequency (DLSF); intra-frame correlation; inter-frame correlation; Split Vector Quantization (SVQ)

0 引言

线性预测(Linear Prediction, LP)技术是语音信号处理中的核心技术,不仅广泛应用于语音编码领域,而且 LP 系数也是语音识别的重要信息来源^[1]。目前 LP 系数最有效的表达方式是与其在数学角度上完全等价的线谱对(Line Spectral Pair, LSP)系数^[2-3]。相对于 LP 系数而言,LSP 系数不仅与语音的短时谱包络密切相关^[4],且拥有更优的量化特性、内插特性和鲁棒性^[5],同时 LSP 系数有良好的动态范围,易保证所构成的滤波器是稳定的,因此 LSP 系数已广泛应用于语音压缩、语音增强、语音分析和语音识别等应用领域中^[6-10]。特别是在低速率语音编码标准中,通常相当多的比特数被用来实现 LSP 系数的量化,因此研究利用 LSP 系数的帧内及帧间相关性和高效率的矢量量化技术来实现进一步的数据压缩具有极其重要的意义。

出于对现代通信信道利用的有效性及经济性的双重考虑^[11],本文从低速率语音编码着手研究 LSP 参数的相关性。本文采用线谱频率(Line Spectrum Frequency, LSF)参数来描

述 LSP 系数的频域值,出于对 LP 谱包络与语音谱包络的拟合程度及计算速度的综合考虑,选择 LSF 参数阶数 p 为 10。本文所研究的 LSF 参数相关性为 LSF 参数的自身属性,并不局限于某一特定语音信号处理领域,旨在更好地理解 LSF 的参数特性,为各应用 LSF 系数的领域提供基础的实验数据支持和参考。

1 语音参数帧内、帧间相关性定义

1.1 语音参数帧内相关性分析与定义

定义 f_n^j 为第 n 帧的第 j 个参数($j = 1, 2, \dots, p; n = 1, 2, \dots, N; N$ 为语音帧总数)。令 $\mathbf{F}^j = \{f_1^j, f_2^j, \dots, f_N^j\}$ 为第 j 维参数,由每帧的第 j 个参数所构成。根据语音的短时平稳特性以及语意的连贯性,假定参数 f_n^j 为离散平稳随机过程,则第 j 维参数 \mathbf{F}^j 的均值 $E(\mathbf{F}^j)$ 和方差 $D(\mathbf{F}^j)$ 为:

$$E(\mathbf{F}^j) = \frac{1}{N} \sum_{n=1}^N f_n^j \quad (1)$$

$$D(\mathbf{F}^j) = \frac{1}{N} \sum_{n=1}^N (f_n^j - E(\mathbf{F}^j))^2 \quad (2)$$

收稿日期:2010-06-23;修回日期:2010-08-12。

作者简介:韩笑蕾(1983-),女,黑龙江哈尔滨人,博士研究生,主要研究方向:语音信号处理、语音压缩编码;赵晓群(1962-),男,黑龙江依安人,教授,博士生导师,博士,主要研究方向:通信与信息理论;方腾龙(1988-),男,安徽枞阳人,硕士研究生,主要研究方向:语音信号处理、语音压缩编码;贾晓光(1981-),男,黑龙江伊春人,硕士,主要研究方向:信息安全、教育信息化。

则第 i 维与第 j 维参数的帧内协方差 $\text{cov}(\mathbf{F}^i, \mathbf{F}^j)$ 和相关系数 $\rho^{i,j}$ 为:

$$\text{cov}(\mathbf{F}^i, \mathbf{F}^j) = \frac{1}{N} \sum_{n=1}^N (f_n^i - E(\mathbf{F}^i))(f_n^j - E(\mathbf{F}^j)) \quad (3)$$

$$\rho^{i,j} = \frac{\text{cov}(\mathbf{F}^i, \mathbf{F}^j)}{\sqrt{D(\mathbf{F}^i)} \sqrt{D(\mathbf{F}^j)}} \quad (4)$$

根据柯西—施瓦茨不等式可知 $|\rho^{i,j}| \leq 1$, 当 \mathbf{F}^i 与 \mathbf{F}^j 的线性关系越强时, $|\rho^{i,j}|$ 就越趋于 1。因此令 $\tilde{\rho} = [\tilde{\rho}^{i,j}] = [|\rho^{i,j}|]$ 为帧内相关系数矩阵 ($i, j = 1, 2, \dots, p$), $\tilde{\rho}^{i,j}$ 表示第 i 维与第 j 维参数的帧内相关系数, $\tilde{\rho}$ 是一个实对称矩阵, 其主对角元素恒为 1, 为 \mathbf{F}^j 的自相关系数, 若各分量之间互不相关, 则 $\tilde{\rho}$ 中只存在主对角元素。

1.2 语音参数帧间相关性分析与定义

设 k 为帧间相隔距离, 定义 $\mathbf{F}^j|_1^m = \{f_1^j, f_2^j, \dots, f_m^j\}$ 为 1 到 m 帧的第 j 个参数所构成的样本 ($m = 1, 2, \dots, N$), 则相距 k 帧的第 j 维参数间的协方差 $\text{cov}(\mathbf{F}^j|_1^{N-k}, \mathbf{F}^j|_{k+1}^N)$ 和相关系数 ρ_k^j 为:

$$\text{cov}(\mathbf{F}^j|_1^{N-k}, \mathbf{F}^j|_{k+1}^N) = \frac{1}{N-k} \sum_{n=1}^{N-k} ((f_n^j - E(\mathbf{F}^j|_1^{N-k}))(f_{n+k}^j - E(\mathbf{F}^j|_{k+1}^N))) \quad (5)$$

$$\rho_k^j = \frac{\text{cov}(\mathbf{F}^j|_1^{N-k}, \mathbf{F}^j|_{k+1}^N)}{\sqrt{D(\mathbf{F}^j|_1^{N-k})} \times \sqrt{D(\mathbf{F}^j|_{k+1}^N)}} \quad (6)$$

令 $\tilde{\rho}_k = [\tilde{\rho}_k^j] = [|\rho_k^j|]$ 为帧间相关系数矩阵 ($j = 1, 2, \dots, p$), $\tilde{\rho}_k^j$ 表示相距 k 帧的第 j 维参数的帧间相关系数。相关系数的大小所表示的意义目前在统计学界尚不一致, 但通常认为相关系数在 0~0.3 范围内呈微相关, 在 0.3~0.5 范围内呈实相关, 在 0.5~0.8 范围内呈显著相关, 在 0.8~1 范围内呈高度相关。

2 LSF 参数相关性分析与矢量量化

为充分研究不同声源的语音特性, 本文分别对英语和汉语的 LSF 参数相关性进行统计和分析。对于英语和汉语分别采用 TIMIT 及 PKU-SRSC 语音数据库, 形成英语女声 (F_1)、英语男声 (M_1)、汉语女声 (F_2) 和汉语男声 (M_2) 4 个语音文件, 持续时间各为 210 s, 并采用帧长 30 ms, 帧移 10 ms 对语音进行分帧处理, 则每个语音文件包含 10507×10 维 LSF 参数。

2.1 LSF 参数帧内相关性分析

根据 1.1 节的定义, 令第 n 帧的 10 个 LSF 系数为 $f_{\text{lsf},n}^1 \sim f_{\text{lsf},n}^{10}$, 通过 Matlab 进行仿真, 分别统计出英语与汉语男女声 LSF 参数帧内相关系数, 结果如图 1 所示, 其中横坐标代表 10 维 LSF 参数, 纵坐标代表各维 LSF 参数间的帧内相关系数。由图 1 可见, 相邻 LSF 参数之间具有很强的帧内相关性, LSF 参数在帧内存在冗余信息。进一步观察发现, 英语男女声 LSF 参数的帧内相关性符合相似统计规律, 汉语男女声 LSF 参数的帧内相关性符合相似统计规律, 因此本文把 F_1 与 M_1 合并成英语语音文件, 把 F_2 与 M_2 合并成汉语语音文件, 并分别计算英语 LSF 参数帧内相关系数 (表 1) 和汉语 LSF 参数帧内相关系数 (表 2)。

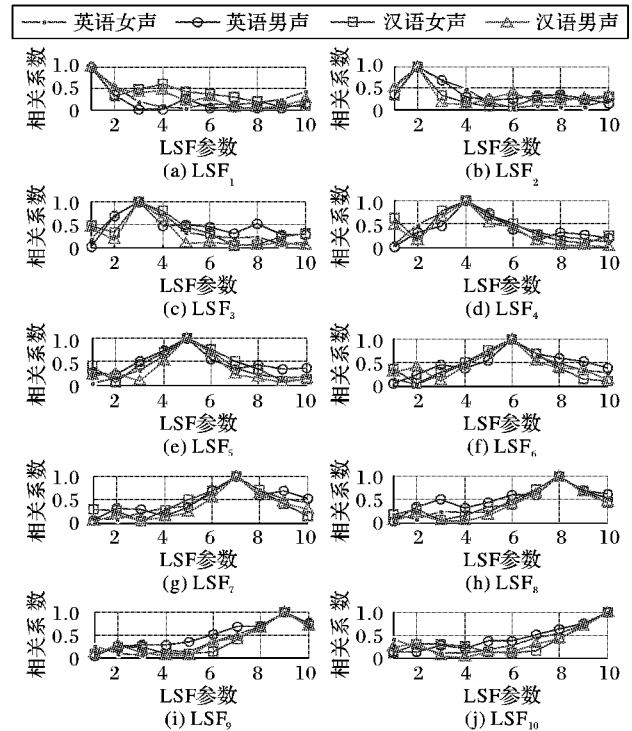


图1 英语与汉语男、女声 LSF 参数帧内相关系数分布图

表1 英语 LSF 参数帧内相关系数 $\tilde{\rho}^{i,j}$ 统计表 ($\times 10^{-3}$)

j	i									
	1	2	3	4	5	6	7	8	9	10
1	1000	445	82	1	156	69	36	84	102	272
2	445	1000	669	380	188	123	182	192	77	66
3	82	669	1000	597	395	342	260	363	170	67
4	1	380	597	1000	643	457	273	303	258	163
5	156	188	395	643	1000	605	310	312	236	265
6	69	123	342	457	605	1000	675	561	455	330
7	36	182	260	273	310	675	1000	547	627	469
8	84	192	363	303	312	561	547	1000	690	570
9	102	77	170	258	236	455	627	690	1000	739
10	272	66	67	163	265	330	469	570	739	1000

表2 汉语 LSF 参数帧内相关系数 $\tilde{\rho}^{i,j}$ 统计表 ($\times 10^{-3}$)

j	i									
	1	2	3	4	5	6	7	8	9	10
1	1000	414	431	544	304	325	189	113	1	72
2	414	1000	227	16	105	236	218	248	258	305
3	431	227	1000	706	263	210	39	64	174	169
4	544	16	706	1000	590	455	229	74	65	145
5	304	105	263	590	1000	717	356	228	56	135
6	325	236	210	455	717	1000	577	365	241	21
7	189	218	39	229	356	577	1000	681	418	215
8	113	248	64	74	228	365	681	1000	661	426
9	1	258	174	65	56	241	418	661	1000	691
10	72	305	169	145	135	21	215	426	691	1000

2.2 LSF 参数帧间相关性分析

根据 1.2 节的定义, 分别统计英语与汉语男女声 LSF 参数帧间相关系数, 结果如图 2 所示, 其中横坐标代表相隔的帧数, 纵坐标代表各维 LSF 参数的帧间相关系数。由图 2 可见, 相邻帧 LSF 参数具有很强的帧间相关性, 即使在相距 2 帧时 $F_{\text{lsf}}^1 \sim F_{\text{lsf}}^2$ 的帧间相关性仍较强, 因此 LSF 参数在帧间存在冗余信息。

2.3 LSF 参数矢量量化

矢量量化比标量量化的效率高, 但是要 LSP 系数进行

精细量化,无论是码书存储量还是计算量和复杂度都是惊人的^[12]。权衡存储量、计算量和硬件的技术水平,目前应用最为广泛的技术是分裂矢量量化(Split Vector Quantization, SVQ)^[13-16]。分裂矢量量化首先把预量化的矢量分裂成几个子矢量,然后对每个子矢量应用独立的码本进行量化。在语音编码中,目前多采用符合人耳听觉特性的对数谱失真测度来衡量量化是否满足透明量化的要求^[17]:1)平均谱失真测度应约为1 dB;2)出现在2~4dB范围内的对数谱失真测度值应不多于2%;3)对数谱失真测度应不大于4 dB。定义第*i*帧的对数谱失真测度 $SD(i)$ 为:

$$SD(i) = \sqrt{\frac{1}{\pi} \int_0^{\pi} \left[10 \lg \frac{S_{w,i}(w, w_i)}{S_{w,i}(w, \hat{w}_i)} \right]^2 dw} \quad (7)$$

其中: $i = 1, 2, \dots, N$, N 为语音帧的总数; $S_{w,i}(w, w_i) = 1/|A_{p,w_i}(e^{jw})|^2$ 为第*i*帧的LPC谱密度; $S_{w,i}(w, \hat{w}_i)$ 为量化后第*i*帧的LPC谱密度。则平均谱失真测度 $Av. SD$ 为:

$$Av. SD = \frac{1}{N} \sum_{i=1}^N SD(i) \quad (8)$$

矢量量化若能有效地利用矢量中各分量间的相关性,将会提高矢量量化质量。由表1与表2可见, F_{lsf}^3 与 F_{lsf}^4 显著相关、 F_{lsf}^5 与 F_{lsf}^6 显著相关、 F_{lsf}^2 和 F_{lsf}^3 同 $F_{lsf}^5 \sim F_{lsf}^{10}$ 之间微相关,因此考虑将 F_{lsf}^3 与 F_{lsf}^4 参数分在同一组内、 F_{lsf}^5 与 F_{lsf}^6 参数分在同一组内,并且不要把 F_{lsf}^2 或 F_{lsf}^3 同 $F_{lsf}^5 \sim F_{lsf}^{10}$ 分为一组。结合归纳出的分组规律与当前应用较广的SVQ分组方法,对英语及汉语LSF参数进行(4,6)、(5,5)、(3,3,4)、(4,2,4)、

(4,4,2)和(4,3,3)六种分组SVQ仿真,并使用谱失真衡量18 bits/帧和20 bits/帧的量化性能,如表3所示。

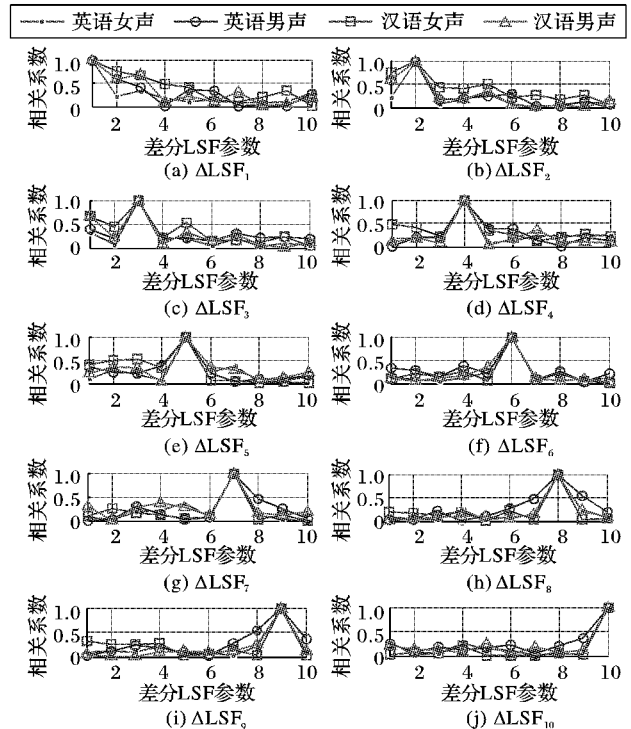


图2 英语与汉语男、女声LSF参数帧间相关性分布图

表3 英语、汉语LSF参数SVQ谱失真

SVQ			英语			汉语		
LSF分组	bits分配	bits/帧	$Av. SD$ /dB	[2,4]dB/%	>4 dB/%	$Av. SD$ /dB	[2,4]dB/%	>4 dB/%
(4,6)	(9,9)	18	0.5741	1.69	0.00	0.7213	3.43	0.01
(5,5)	(9,9)	18	0.6225	2.79	0.01	0.8224	5.56	0.03
(3,3,4)	(6,6,6)	18	0.6584	2.69	0.01	0.8267	5.21	0.01
(4,2,4)	(8,3,7)	18	0.6642	2.56	0.00	0.8174	5.08	0.05
(4,2,4)	(8,4,6)	18	0.6392	2.51	0.00	0.8023	4.76	0.03
(4,4,2)	(7,7,4)	18	0.6793	3.09	0.01	0.8625	6.18	0.02
(4,4,2)	(8,7,3)	18	0.6366	2.25	0.00	0.7949	4.66	0.02
(4,3,3)	(7,6,5)	18	0.6865	3.19	0.00	0.8590	5.94	0.03
(4,6)	(10,10)	20	0.5120	1.33	0.00	0.6449	2.20	0.01
(5,5)	(10,10)	20	0.5648	2.10	0.00	0.7493	3.98	0.02
(3,3,4)	(7,6,7)	20	0.5912	2.14	0.00	0.7549	3.84	0.01
(4,2,4)	(9,4,7)	20	0.5875	2.01	0.00	0.7293	3.41	0.01
(4,4,2)	(9,8,3)	20	0.5775	1.80	0.00	0.7189	3.37	0.00
(4,3,3)	(8,7,5)	20	0.6135	2.27	0.00	0.7857	4.58	0.05

基于计算量及计算机运算速度的考虑,选取时长为420 s的英语及汉语语音文件来训练SVQ量化器,虽然训练样本数相对有限,但是由表3仍可知:1)(3,3,4)和(5,5)分组SVQ在18 bits/帧及20 bits/帧情况下均没有达到透明量化;2)无论是英语还是汉语,(4,6)分组的SVQ量化性能明显优于其他分组方案,且英语在18 bits/帧即达到透明量化;3)在20 bits/帧时,英语(4,6)及(4,4,2)分组SVQ均满足透明量化,(4,2,4)分组SVQ接近于透明量化,且(4,4,2)及(4,2,4)分组的最大子码书容量仅为(4,6)分组最大子码书容量的一半;4)通过比较(4,2,4)与(4,4,2)在18 bits/帧情况下两种编码比特数分配方案的量化结果发现,根据人耳听觉特性来分配量化的比特数,即对敏感频率段对应的LSF参数分配较多的比特数,对不敏感频率段对应的LSF参数分配较少的比特数,会得到性能更优的量化器;5)使用SVQ方法对LSF参

数进行矢量量化时,其量化性能同分组方式及子码书的大小有关,且每帧所分配的比特数越大,量化性能越好。因此在使用SVQ对LSF参数进行量化时,若在相同编码速率下使用(4,6)分组代替G.723.1编码标准中的(3,3,4)分组^[18]及G.729编码标准中的(5,5)分组^[19],将可改善编码性能,提高合成语音的质量。并且在对码书容量没有要求的情况下,采用(4,6)分组会得到较优的量化器;否则采用(4,2,4)或(4,4,2)分组会得较优的量化器。

3 差分LSF参数相关性分析与量化

3.1 差分LSF参数帧内相关性分析

LSF参数在帧内及帧间都具有很强的相关性,LSF参数序列中存在大量冗余信息,难以实现高质量、低速率的语音编码,因此提出了差分LSF(Differential LSF, DLSF)参数的概念。

由于各维 DLSF 参数动态范围更小,因此具有比 LSF 参数更好的量化特性^[20-21]。令 10 阶 DLSF 参数为 ΔLSF :

$$\Delta LSF = \{\Delta f_{lsf,n}^1, \Delta f_{lsf,n}^2, \dots, \Delta f_{lsf,n}^{10}\} = \{f_{lsf,n}^1, f_{lsf,n}^2 - f_{lsf,n}^1, \dots, f_{lsf,n}^{10} - f_{lsf,n}^9\} \quad (9)$$

仍用 F_1 、 M_1 、 F_2 和 M_2 共 4 个语音文件统计英语与汉语男女声 DLSF 参数帧间相关性,结果如图 3 所示。比较图 1 和图 3 发现各维 DLSF 参数的帧内相关性较 LSF 参数的帧内相关性小。

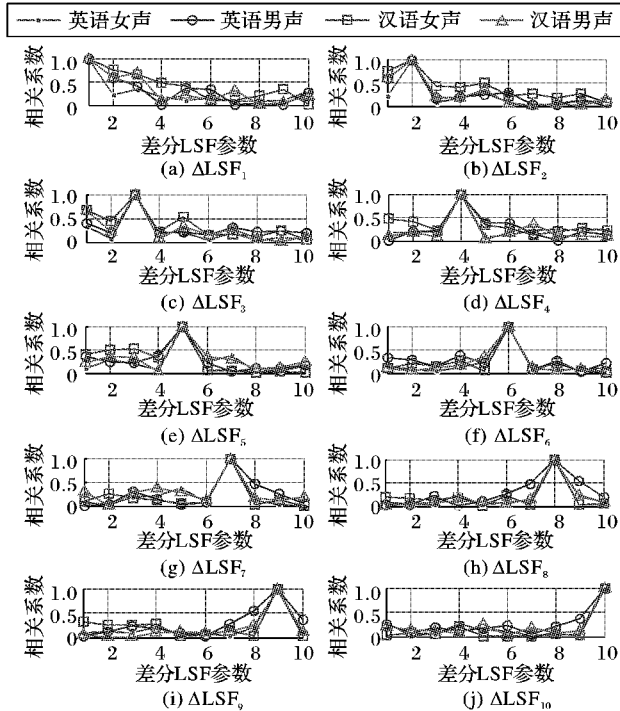


图3 英语与汉语男、女声 ΔLSF 参数帧内相关性分布图

本文按 1.2 节中的定义归纳英语与汉语男女声 LSF 参数与 DLSF 参数的帧内相关程度分布情况,如表 4 所示。由于帧内相关系数矩阵 $\hat{\rho}$ 中主对角线上的元素均为 1,因此表 4 中各阶参数同时存在 10% 的高度相关。由表 4 可见英语中至少有 68% 的 DLSF 参数在帧内呈微相关,汉语中至少有 70% 的 DLSF 参数在帧内呈微相关,DLSF 参数有效地减少了

LSF 参数在帧内存在的冗余信息。

表4 LSF 参数与 DLSF 参数帧内相关程度分布表 %

相关程度	英语女声		英语男声		汉语女声		汉语男声	
	LSF	DLSF	LSF	DLSF	LSF	DLSF	LSF	DLSF
高度相关	10	10	10	10	10	10	10	10
显著相关	22	0	20	4	18	8	12	4
实相关	48	4	26	18	26	12	22	10
微相关	50	86	44	68	46	70	56	76

3.2 差分 LSF 参数帧间相关性分析

按照 1.2 节的定义统计英语与汉语男女声 DLSF 参数帧间相关性,计算结果与图 2 相似,差分 LSF 参数间存在帧间相关性,且相邻帧间各维 DLSF 参数呈显著相关或高度相关,因此 DLSF 参数在帧内仍存在冗余信息。

3.3 差分 LSF 参数矢量量化

由图 3 知各维 DLSF 参数间的帧内相关程度较弱,因此考虑对各维 DLSF 参数进行标量量化。令量化后的第 n 帧 10 阶差分 LSF 参数为 $\Delta \hat{f}_{lsf,n}^1, \Delta \hat{f}_{lsf,n}^2, \dots, \Delta \hat{f}_{lsf,n}^{10}$, 量化误差为 $\varepsilon_n^1, \varepsilon_n^2, \dots, \varepsilon_n^{10}$, 则通过式(9)可推导出:

$$\{f_{lsf,n}^1, f_{lsf,n}^2, \dots, f_{lsf,n}^{10}\} = \{\Delta f_{lsf,n}^1, \Delta f_{lsf,n}^2 + \Delta f_{lsf,n}^1, \dots, \Delta f_{lsf,n}^{10} + \Delta f_{lsf,n}^9 + \dots + \Delta f_{lsf,n}^1\} = \{\Delta \hat{f}_{lsf,n}^1 + \varepsilon_n^1, \Delta \hat{f}_{lsf,n}^2 + \Delta \hat{f}_{lsf,n}^1 + \varepsilon_n^1 + \varepsilon_n^2, \dots, \Delta \hat{f}_{lsf,n}^{10} + \Delta \hat{f}_{lsf,n}^9 + \dots + \Delta \hat{f}_{lsf,n}^1 + \varepsilon_n^1 + \dots + \varepsilon_n^{10}\} \quad (10)$$

观察式(10)发现若对 DLSF 参数采用与 LSF 参数相同的量化方法量化,则其量化误差会逐维传递、叠加,因此本文采用文献[20]中的 DSQ 方法,并在 18 bits/帧时按(3,2,2,2,2,2,2,1,1,1)为 10 阶 DLSF 参数进行比特分配;在 20 bits/帧时按(3,3,3,2,2,2,2,1,1,1)进行比特分配,其谱失真结果如表 5 所示。为与 LSF 参数的量化性能相比较,表 5 中同时列出了相同编码条件下 LSF 参数的标量量化谱失真结果。由表 5 可见,相对于 LSF 参数而言,DLSF 参数的量化失真较小,有效抑制了易被人耳所感知的大于 4 dB 的谱失真,同时在 20 bits/帧情况下,英文 DLSF 参数的 DSQ 接近透明量化,且量化质量随每帧分配的比特数增加而提高,因此在相同编码速率下,DLSF 参数具有比 LSF 参数更优良的标量量化特性。

表5 英语、汉语 DLSF 参数与 LSF 参数标量量化谱失真

参数	Bits/帧	英语			汉语		
		Av. SD/dB	[2,4] dB/%	>4 dB/%	Av. SD/dB	[2,4] dB/%	>4 dB/%
DLSF	18	0.6947	2.59	0.01	0.9028	6.49	0.02
	20	0.6332	1.67	0.01	0.7685	4.04	0.02
LSF	18	0.7934	3.47	0.05	0.9082	6.26	0.08
	20	0.6820	2.09	0.02	0.794	4.71	0.04

根据香农信息论,矢量量化能有效地利用矢量中各分量间的相关性,因此矢量量化优于标量量化。由于矢量量化的码书设计,是把 M 个训练矢量看做一个整体,并按照失真测度最小的准则,尽可能地搜索到全局最优或接近全局最优的含 N 个码字的码书^[22],在码书训练过程中的每一次迭代搜索都是寻求一组矢量的多维模式匹配、优化的过程,因此本文采用 SVQ 方法训练 DLSF 码书,其量化误差并不会按照式(10)在各维参数间简单传递、叠加,同时采用可消除量化误差传递、叠加的 EEDSVQ 算法对 DLSF 参数进行量化^[23]。由图 3 可见,前 3 维 DLSF 参数之间的帧内相关程度较强,因此本文在 2.3 节中所采用的 6 种分组方案的基础上加入了(3,7)分

组,并在 18 bits/帧时按(7,11)分配比特,在 20 bits/帧时按(8,12)分配比特。

通过观察仿真结果发现:

1) 汉语 DLSF 参数的量化失真略小于 LSF 参数的量化失真,而英文 DLSF 参数的量化失真略大于 LSF 参数的量化失真,造成这一现象的原因有二:一是本文用以训练 DLSF 码书的英语及汉语的样本数相对有限,使得量化器没有达到最优;二是英语与汉语属于不同语种,其 DLSF 参数之间存在较大差异,从而导致汉语与英语的 EEDSVQ 量化器产生了不同的量化效果。语种的不同对 DLSF 参数量化器所产生的具体影响,本文没有进一步探究。

2)通过比较(4,2,4)与(4,4,2)在18 bits/帧情况下两种编码比特数分配方法的量化结果发现,根据人耳听觉特性来分配 DLSF 参数量化的比特数,会得到性能更优的量化器。

3)英文 DLSF 参数(3,7)分组在18 bits/帧即达到透明量化,且量化性能优于 LSF 参数在相同编码速率下的(4,6)分组。

4)即使 DLSF 参数的帧内相关性非常小,在相同编码速率下对 DLSF 参数所采用的7种分组方案的 EEDSVQ 量化失真全部小于采用 DSQ 的量化失真,且随着每子帧所分配的比特数越大,量化性能越好。结合表5,发现在相同编码速率下,DLSF 参数的量化性能优于 LSF 参数。

4 结语

由于 LSF 参数同时具有帧内及帧间相关性,因此现代语音编码都致力于研究如何利用这些相关性来实现高质量、低速率的语音编码。本文对英语及汉语男女声分别进行了 LSF 参数的帧内及帧间相关性统计,并结合统计结果对英语及汉语的 LSF 参数进行了6种分组的 SVQ 仿真。通过实验结果观察到:1)在对 LSF 参数采用 SVQ 量化时,当对码书大小没有要求的情况下,采用(4,6)分组量化性能较好;否则,采用(4,2,4)或(4,4,2)分组的量化性能较好。2)在相同编码速率下,对 LSF 参数采用(4,6)分组替换 G. 732.1 标准中的(3,3,4)分组及 G. 729 标准中的(5,5)分组,可改善 G. 723.1 及 G. 729 标准的编码质量。随后对 DLSF 参数进行相关性统计,发现至少68%的 DLSF 参数在帧内呈微相关,大大降低了 LSF 参数的帧内冗余度。随后结合 Δ LSF 参数的帧内相关性,采用 DSQ 及七种分组的 EEDSVQ 方法对 DLSF 参数进行量化,实验结果表明在相同编码速率下,DLSF 参数的量化性能优于 LSF 参数。

文中所示的量化器谱失真均采用 F_1, M_1, F_2, M_2 4个实验语音文件的相干语音进行统计,当采用实验语音文件外的非相干语音进行统计时,其谱失真比相干语音量化后的谱失真平均约大0.4 dB,但是由分组方法及每帧所分配比特数的不同而导致的量化器质量的差异,同相干语音符合相同规律。无论是 LSF 参数还是 DLSF 参数的量化器性能,均同分组方案及子码书的大小密切相关,根据 LSF 参数及 DLSF 参数的帧内相关性统计结果,可制定较为合理的分组方法,排除一些不可能的组合方案。

尽管本文在训练 SVQ 及 EEDSVQ 量化器时所采用的样本数相对略少,但从实验结果仍可以看出不同分组方法对 SVQ 及 EEDSVQ 量化器性能影响的趋势。下一步的工作,将选取实验最优的分组方案,大量增加样本数,训练出性能较优的 EEDSVQ 量化器,并利用 LSF 参数和 DLSF 参数同时存在的帧间相关性,实现变帧长的帧间预测方法,从而大幅度降低参数量化时所需比特数。

参考文献:

- [1] 胡航. 语音信号处理[M]. 哈尔滨: 哈尔滨工业大学出版社, 2000.
- [2] ITAKURA F. Line spectrum representation of linear predictive coefficients of speech signals [J]. Journal of the Acoustical society of America, 1975, 57(S1): 35.
- [3] SUGAMURA N, ITAKURA F. Speech analysis and synthesis methods developed at ECL in NTT — from LPC to LSP [J]. Speech Communication, 1986, 5(2): 213 – 229.
- [4] MCLOUPHILIN I V. Review: Line spectral pairs [J]. Signal Processing, 2008, 88(3): 448 – 467.
- [5] SAITO S, NAKATA K. Fundamentals of speech signal processing [M]. New York: Academic Press, 1985.
- [6] de ALENCAR V F S, ALCAIM A. Digital filter interpolation of decoded LSFs for distributed continuous speech recognition [J]. Electronics Letters, 2008, 44(17): 1039 – 1040.
- [7] SAHIDULLAH M, SAHA G. On the use of perceptual line spectral pairs frequencies for speaker identification [C]// The National Conference on Communications. Washington, DC: IEEE, 2010: 1 – 5.
- [8] KIM H K, KIM K C, LEE H S. Enhanced distance measure for LSP-based speech recognition [J]. Electronics Letters, 1993, 29(16): 1463 – 1465.
- [9] HASEGAWA-JOHNSON M. Line spectral frequencies are poles and zeros of the glottal driving-point impedance of a discrete matched-impedance vocal tract model [J]. Acoustical Society of America, 2000, 108: 457 – 460.
- [10] MORRIS R W, CLEMENTS M A. Modification of formants in the line spectrum domain [J]. IEEE Signal Processing Letters, 2002, 9(1): 19 – 21.
- [11] 赵晓群. 数字语音编码[M]. 北京: 机械工业出版社, 2007.
- [12] CAMPBELL J P, WELCH V C, Jr, TREMAIN T E. An expandable error-protected 4800 bps CELP coder (US Federal Standard 4800 bps voice coder) [J]. Acoustics, Speech, and Signal Processing, 1989, 2: 735 – 738.
- [13] PALIWAL K K, ATAL B S. Efficient vector quantization of LPC parameters at 24 bits/frame [C]// IEEE International Conference on Acoustic, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1991: 661 – 664.
- [14] CHEN J H, WANG D. Transform predictive coding of wideband speech signals [C]// IEEE International Conference on Acoustic, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1996, 1: 275 – 278.
- [15] CHANG D I, CHO Y K, ANN S. Efficient quantization of LSF parameters using classified SVQ combined with conditional splitting [C]// IEEE International Conference on Acoustic, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1995, 1: 736 – 739.
- [16] SUNG H, SON C Y, LEE K, et al. An efficient LSF quantization using dynamic bit allocation [C]// IEEE International Conference on Acoustic, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 2009: 4141 – 4144.
- [17] CHANG D I, ANN S, LEE C W. A classified split vector quantization of LFS parameters [J]. Signal Processing, 1997, 59(3): 267 – 273.
- [18] ITU-T. Recommendation G. 723.1: dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 Kbit/s [S]. ITU-T, 1996.
- [19] ITU-T. Recommendation G. 729: coding of speech at 8 Kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP) [S]. Geneva: ITU-T, 1996.
- [20] MERAZKA F. LSF quantization to enhance the frame erasure. Robustness of CELP type coders in packet [C]// World Congress on Engineering 2009. London: IAENG, 2009: 230 – 233.
- [21] MERAZKA F. Enhanced differential split vector quantization of line spectrum pairs for CELP-type coders in packet networks [C]// World Congress on Engineering and Computer Science 2009. San Francisco: IAENG, 2009: 1326 – 1329.
- [22] 孙圣和, 陆哲明. 矢量量化技术及应用[M]. 北京: 科学出版社, 2002.
- [23] HAN X L, ZHAO X Q, FANG T L, et al. An efficient differential LSFs quantization for chinese mandarin speech [C]// International Conference on Communication Systems. Hong Kong: CPS, 2010: 263 – 266.