

## NAQ 与韵律特征的关系初探

高莹莹, 朱维彬

(1. 北京交通大学 信息科学研究所, 北京 100044, 中国;

2. 北京交通大学, 北京 100044, 中国)

**摘要:** 通过逆滤波方法实现了连续语流的声门波提取, 并利用音节能量质点处的 NAQ (Normalized Amplitude Quotient) 来代替音节的总体 NAQ 从而使其更稳健, 参数表现更简洁。然后, 分析了 NAQ 与重音、韵律层级结构等韵律特征的关联关系。结果显示: 在单焦点情形下, 重音词内每个音节 NAQ 都有明显抬升, 且后面音节 NAQ 大于前面音节; 重音词前后的非重音成分 NAQ 受到抑制, 且这种“强-弱”对比在用 NAQ 表现时更加明显。在双焦点和多焦点环境下, 用 NAQ 同样能体现出重音层级性, 核心重音都在最后一个重音上, 但表现不出核心前重音。另外发现单一质点上的 NAQ 与韵律层级结构没有一种明确的表现关系。

**关键词:** NAQ; 重音; 韵律层级结构

**中图分类号:** TP 3

### The relationship between NAQ and prosodic features

Gao Yingying, Zhu Weibin

(1. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

2. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** We first improve the inverse filter algorithm and successfully extract the glottal wave of continuous speech. Use NAQ (Normalized Amplitude Quotient) at the energy centroid of the syllable to stand for the whole NAQ of the syllable, which make it steady and simpler. Then we analysis the relationship between NAQ and some prosodic features such as stress and prosodic hierarchy. We find that under the condition of single focus, each syllable in the stress word has a significant uplift on NAQ, and the NAQ of the after syllable is larger than the previous one. The NAQ of the non-stress compositions is inhibited, and this "strong - weak" relationship is more obvious with the usage of NAQ. And because of the single parameter, simple control conditions, it is easier to build a stress synthesizer. With the using of NAQ to perform stress, the stress hierarchy also is presented under the condition of double and multiple focuses, the core stress is the last stress, but it dose not present a former core stress, which is different with the performance when performing stress with pitch[7].

**Key words:** NAQ; stress; prosodic hierarchy;

## 1 引言

随着语音合成技术向更高层次发展, 人们越来越关注激励源部分的提取与应用, 传统的分析-合成算法基本仅用基频来简单模拟激励源, 合成效果非常有限, 尤其是发音风格的模拟上, 与人的声音还有一定差距。

对于音色的定量分析主要依据声门激励的特性。最早提取声门波的方法是通过声门仪采集声门开启和关闭时的生理电信号, 逆滤波的引进为我们提供了一种非入侵的方法来估计声门激励。在提取到声门波后, 还需要对其进行量化或模型描述, 从而更好的应用在工程上。有很多对声门激励的量化参数, 常用的时域参数有开启商(open quotient, OQ)、关闭商(closing quotient, CQ)和速度商(speed quotient, SQ), 其中以关闭商最为常用, 原因有两点: (1) 物理意义上, 关闭商即关闭相占整个基音周期的比例, 它直接体现声源谱幅度的倾斜程度, 而这又影响了高频能量的分布, 从而体现发音方式的变化[1] (2) 实现手段上, 这三种商直接求都需要提取时刻值, 这样很难保证结果的精确性, 但关闭商可以通过幅度参数来近似, 即归一化幅度商(normalized amplitude quotient, NAQ), 它是声门波及其导数的幅度商归一化之后的结果。实验证明: NAQ 比时域参数 CQ 对语音噪声和失真更鲁棒[2], 并对发音人的身份、行为[3]和发音方式[2] [3]等都有明显的区分性。

Nick Campbell[3][4], Paavo Alku[2][5] 以及国内的白洁[6]等人都做过提取 NAQ 这方面的工作, 本文在这些工作的基础上做了些改进, 对怎样提取

**基金项目:** 校科研基金(KKJ07006532)

**作者简介:** 高莹莹(1987), 女(汉), 河北, 博士生在读

**通讯作者:** 朱维彬, 副教授, E-mail: wbzhu@bjtu.edu.cn

到更加稳健的 NAQ 做了探索。为了将 NAQ 更好的应用到以后工作中，本文首先从基础层面分析了 NAQ 与重音、韵律层级结构等韵律特征的关联关系。

## 2 NAQ 的提取

首先利用迭代自适应逆滤波（iterative adaptive inverse filter, IAIF）提取声门波，然后计算音节的能量质点，提取能量质点处的声门波及其导数的幅值计算 NAQ，将其作为音节的总体 NAQ，参数更加简洁，也更稳定，可信度更高。

### 2.1 提取声门波

IAIF 是一种从语音信号中提取声门激励的分析方法。其基本原理是：通过循环迭代地利用逆滤波技术，尽可能从原始语音信号中去除声门激励和口鼻辐射的效应，以便更精确地估计声道滤波器特性，从而尽可能从语音信号中去除声道滤波器效应，最终更精确地估计声门激励信号。IAIF 流程如图 1 所示。Paavo Alku[5]、白洁[6]等人在实现时都用 DAP（Discrete All Pole）来代替 LPC 进行模型估计，从我们的实验结果来看，用 LPC 就能得到比较理想的声门波，还大大降低了计算复杂度。

可以看出整个流程可以虚线为界分为两个阶段：第一阶段通过较低阶次的 LPC 估计、逆滤波、积分等操作得到声门波的初步估计；第二阶段在此基础上再利用 LPC 估计等上述步骤得到更为准确的声门波。

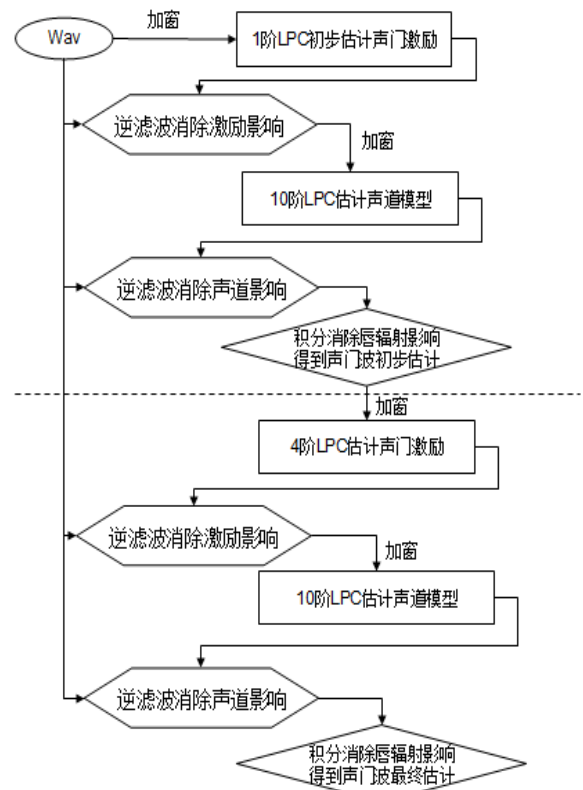


图 1 IAIF 提取声门波流程图

具体实施需注意以下细节：（1）逆滤波的输入是未加窗的原始语音数据，一般取一个基音周期长度；（2）每次 LPC 估计激励源或声道模型之前都需进行加窗的补偿操作，因为逆滤波的输出也是未加窗的数据；（3）为消除逆滤波器在数据开始位置效果不稳定的情况，最后一次逆滤波的输入数据可分别向前后延展一个基音周期长度，再通过加基音同步窗并叠加的操作合成该基音周期的声门波，这样得到的结果与前后两基音周期的声门波也有关，一定程度上保证了声门开启与闭合的渐变性。

图 2 是一个提取实例，最上面一行是语音波形，中间是声门波导数，最下面是声门波。可以看出提取效果比较理想，共振峰成分已经基本滤除。

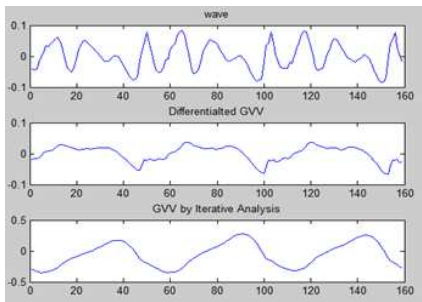


图 2 IAIF 提取的声门波（glottal volume velocity, GVV）

## 2.2 提取音节能量质点处的 NAQ

为了提取到稳定的 NAQ, 并使其更加简洁并具有代表性, 我们选择音节能量质点处的 NAQ 来代表音节的总体 NAQ, 音节的能量质点  $k$  采用以下计算式:

$$k = \sum_{n=1}^N n \cdot s^2(n) / \sum_{n=1}^N s^2(n) \quad (1)$$

振幅商(amplitude quotient, AQ)是用单一的数字值定量描述声源特征最有效的参数之一。它被定义为声门波最大振幅与其相应一阶导数的最大负峰值之比。

$$AQ = f_{ac} / d_{peak} \quad (2)$$

其中:  $f_{ac}$  是声门脉冲的最大波峰值;  $d_{peak}$  是声门脉冲对应一阶导数的最大负峰值。

因为无须测量声门波开启或闭合的瞬间时刻, AQ 值比较容易得到。由于 AQ 的值依赖于信号的基频(F0), 将 AQ 用基音周期归一化, 即导出一个新的声源时域参数 NAQ, 去除了这种对基频的依赖性。

$$NAQ = AQ / T = f_{ac} / (d_{peak} \times T) \quad (3)$$

其中  $T$  为基音周期。

## 3 NAQ 与重音的关系

对于重音的声学表现, 已有的工作通常都用音高和时长来分析, 我们通过观察语句的 NAQ 曲线发现 NAQ 随重音的分布也有明显的变化, 因此我们希望找到这一表现关系, 并与贾媛从音高角度分析重音的实验[7]做一对比, 看 NAQ 在重音表现上是否与音高相同。

实验数据: 不同焦点环境下的平行语料, 发音内容为“麦丽半夜虐待路娜了”。共 48 句, 分宽焦点、单焦点(分主语、宾语两种)、双焦点(主语+宾语)、多焦点(分主语+状语+谓语、状语+谓语+主语两种)六组, 每组 8 句, 4 个发音人, 2 男 2 女, 每人 2 句, 每组内发音内容及焦点环境均相同, 各组发音人均相同。

### 3.1 NAQ 与重音的关系——单焦点环境(排除

### 重音层级的影响)

分别作了主语焦点、宾语焦点和宽焦点的 NAQ 曲线。结果如图 3 所示, 纵坐标是每个音节质心处 NAQ 的平均值。

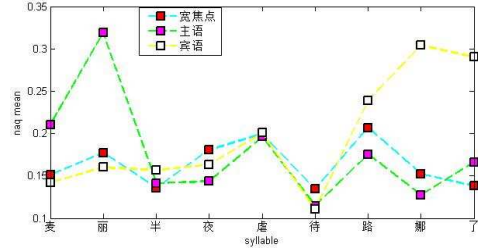


图 3 宽焦点和单焦点 NAQ 曲线图

由图 3 可以看出:

1) 在宽焦点环境下, 核心重音位于句末的句法成分上; 在单焦点时, 焦点与重音一一对应, 负载焦点的成分即负载重音的成分。这点与贾媛用音高得出的实验结论相同[7]。

2) 重音词内每个音节 NAQ 都有明显抬升, 且后面音节 NAQ 大于前面音节, 说明重音词内部也有一个“相对强弱”的节律关系, 此处均表现为“弱-强”关系;

3) 重音词前后的非重音成分 NAQ 受抑制, 这也可以用节律中的“强-弱 (S-W)”关系来解释: 当窄焦点后面有后接的句法成分时, 其节律关系为“强-弱 (S-W)”, 因此, 重音后面的成分的 NAQ 被压低, 且这种作用可以延续到负载句的末尾, 此时, “S”对应焦点位置, “W”对应焦点后位置所有句法成分; 当焦点成分后面没有后接的句法成分时, 其节律结构为“弱-强 (W-S)”, “S”依然对应焦点位置的句法成分, 而“W”对应焦点前的所有句法成分[7]。

4) 用 NAQ 表现的重音词与非重音成分的这种“强-弱”对比关系比用音高表现时更加明显, 这对于构建重音合成器十分有利, 而且目前用到的仅是 NAQ 一个单一参数, 比以往的重音合成器[8]需要用到音阶、音高斜率和时长等多重参数要简便很多, 控制条件上也不用再考虑声调组合等繁复的影响因素。

### 3.2 由 NAQ 表现的重音层级关系——双焦点和多焦点环境

由双焦点和多焦点 NAQ 曲线图可以看出，当用 NAQ 作为重音的声学表现时，同样能体现双焦点和多焦点环境下的重音层级性，并可以得出与贾媛的音高实验[7]相同的结论，即核心重音在最后一个焦点词上；不同的是，双焦点情况下，焦点分布与重音不再成对应关系，而是只实现了最右侧的核心重音，表现不出核心前重音。另外实现方式上核心重音的 NAQ 明显小于单焦点时重音词的 NAQ。

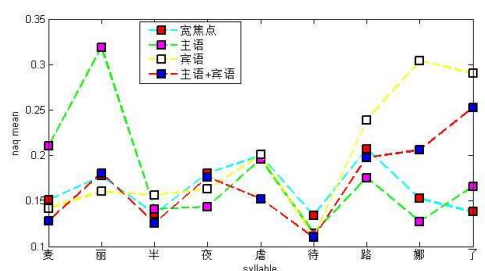


图 4 双焦点与单焦点时 NAQ 对比图

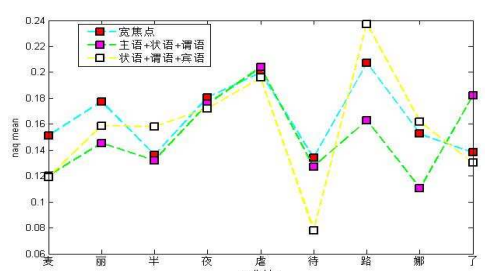


图 5 多焦点 NAQ 曲线图

综上，我们得出：在单焦点环境时，重音词内每个音节 NAQ 都有明显抬升，且后面音节 NAQ 大于前面音节；重音词前后的非重音成分 NAQ 受抑制；用 NAQ 更能表现重音词与非重音成分的“强-弱”对比关系，且参数单一，控制条件简单，更容易构建重音合成器。在双焦点和多焦点环境下，用 NAQ 同样能体现出重音层级性，核心重音都在最后一个重音上，但表现不出核心前重音，这与用音高表现重音时的效果[7]不同。从节律关系来分析，NAQ 更能突显核心重音，使之更强，而抑制其他非核心重音，使之变弱。

#### 4 NAQ 与韵律层级结构的关系

实验语料：宽焦点语料，发音内容仍然为“麦丽 半夜 虐待 露娜 了”，全部为去声，共 12 句，6 个发音人，每人 2 句。

因为句长较短，基本相当于一个短语，所有

只统计了韵律词内及韵律词间的 NAQ 变化，包含四个双音节韵律词和一个单音节词，在考察韵律词内首尾音节的 NAQ 高低变化时不包含单音节词，考察韵律词间即韵律词边界前后音节的 NAQ 变化时包含该单音节词。表 1 分别统计了全部音节的 NAQ 分布范围，韵律词边界前后音节的平均 NAQ，韵律词内首尾音节的 NAQ，并用 T-检验评估了这两组均值的差异性是否显著。

表 1 韵律层级结构的 NAQ 统计

全部音节		
均值	方差	范围
0.1662	0.0074	0.0380-0.4555
韵律词边界		
前	后	T-test
0.1636	0.1757	equal
韵律词		
首	尾	T-test
0.1731	0.1636	equal

从全部音节的 NAQ 分布范围来看，跨度较大，参考 Alku 等人在文献[2]中得到的结论，该实验语料分别涵盖了 *breathy*, *normal* 和 *pressed* 几种发音风格，这可能也是造成后面韵律词边界及韵律词内音节 NAQ 变化不显著的原因之一。

上面提到，从 T-检验结果来看，当前实验条件下，韵律词边界及韵律词内音节的 NAQ 的均值变化是不显著的，即在平均意义上，韵律词内和韵律词间的音节的 NAQ 基本保持稳定，并没有随韵律层级结构的变化而有明显变化。而从韵律词内及韵律词间 NAQ 相对变化分布图（图 6 和图 7）上也可以看出，韵律词内和韵律词间的 NAQ 既有变高也有变低，且高低比例相当，因此我们不能确定 NAQ 随韵律层级结构的变化有一种明显的变化趋势。

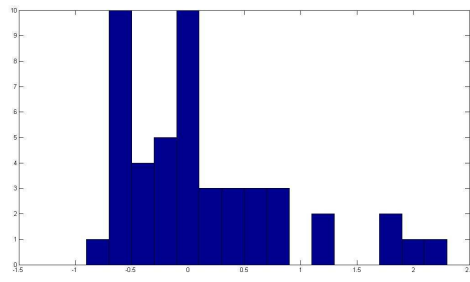


图 6 韵律词内首尾音节的 NAQ 相对变化分布



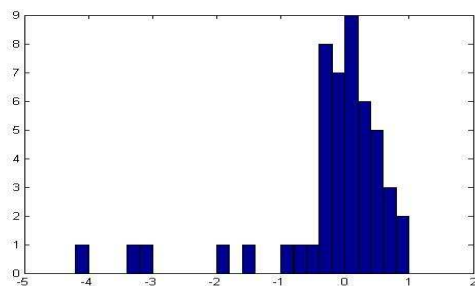


图 7 韵律词边界前后音节的 NAQ 相对变化分布

我们已经知道 NAQ 是描述发音方式的重要参数，而我们希望进一步找到这种描述关系中具体的韵律表现，从现在实验结果来看，这种表现并不明显。分析其原因，首先从实验的完备性上来说，我们仅有“去声+去声”这一种声调组合，其他声调组合还都未涉及，我们不能排除其他组合条件下 NAQ 的韵律表现会更加明显；其次从实验的严谨性上来看，我们仅排除了重音变化对 NAQ 的影响，而没考虑发音风格上的控制，不同发音人乃至同一发音人的不同时刻都可能有不同的发音风格，从前面分析的 NAQ 的分布范围来看，该实验语料包含了 *breathy*, *normal* 和 *pressed* 几种不同发音风格，因此在韵律体现上趋于平均而没有一种明显的关系；另外从目前参数的使用来看，我们用到的仅是单一质心点上的 NAQ，不能排除韵律层级结构与音节内部 NAQ 包络的变化有关，因此我们下一步计划扩展的 NAQ 参数的维度，加入描述 NAQ 包络信

息的参数，以期发现 NAQ 与韵律层级结构更为明确的表现关系。

## 参 考 文 献 (Reference)

- [1] 吴宗济, 林茂灿, 等. 实验语音学概要[H]. 高等教育出版社, P82-83  
Wu Zongjin, Lin Maocan, et al. Summary of Experimental Phonetics Higher Education Press P82-83. (in Chinese).
- [2] P. Alku, T. Bäckström, E. Vilkman .Normalized amplitude quotient for parametrization of the glottal flow. J. Acoust. Soc. Am., vol.112, no.2, pp.701-710, 2002
- [3] N CAMPBELL, MOKHTARI . Voice quality: the 4th prosodic dimension, in 15th ICPhS (ICPhS'03), Barcelona, Spain, 2417-2420 .2003
- [4] P. Mokhtari, N. Campbell. Automatic Measurement of Pressed/Breathy Phonation at Acoustic Centres of Reliability in Continuous Speech. Trans IEICE Special Issue on Speech Information Processing, March 2003.
- [5] P Alku, E Vilkman . Estimation of the glottal pulseform based on discrete all-pole modeling. The 3rd International Conference on Spoken Language 1994 - ISCA
- [6] 白洁, 蒋冬梅, 等 基于 NAQ 的语音情感识别研究 计算机应用研究, 2008 年 第 5 卷第 11 期  
Bai Jie , Jiang Dongmei et al. Emotion recognition based on normalized amplitude quotient . Application Research of Computers V01. 25 No. 11 NOV. 2008 (in Chinese)
- [7] 贾媛、李爱军 论普通话重音的层级性——基于语音事实的分析 第九届中国语音学学术会议论文集, 2010  
Jia Yuan , Li Aijun. The Hierarchy of stress in Mandarin – the Analysis based on the Fact of Speech. PCC 2010 (in Chinese)
- [8] 朱维彬 支持重音合成的汉语语音合成系统 中文信息学报 2007 年 21 卷 03 期  
Zhu Weibin .Mandarin Speech Synthesis System Supporting Stress Synthesis.Journal of Chinese Information . V01 . 21 No.3 NOV. 2007 (in Chinese)