The Chinese Membrane Flute (*dizi*): Physics and Perception of its Tones

Chen-Gia Tsai

Abstract

The goal of this thesis is to address the acoustic and psychological foundations of the music of the Chinese membrane flute (*dizi*). The *dizi* has a membrane covering a hole in the wall of the instrument between the mouth hole and the uppermost finger-hole. The membrane can enhance upper harmonics of *dizi* tones, but it also restricts the tone range. Reductions in the playability of notes in the third octave were explained by input impedance calculations and measurements. Wrinkles in the membrane are critical to the production of the characteristic *dizi* timbre. Theoretical and experimental studies showed that wrinkling can reduce the membrane¡¦s cubic non-linearity, thereby avoiding the jump phenomenon of a Duffing oscillator.

Another effect of the membrane is the inhomogeneity of the *dizi* timbre. As the membrane is driven by the acoustic pressure in the pipe and radiates sounds, it vibrates more when pressure antinodes lie below it. According to spectral features, the *dizi* tone range can be divided into five registers. Three major spectral features, (1) the first formant at 4¡V6 kHz, (2) the predominance of odd-numbered harmonics, and (3) high-frequency subharmonics $(2n\text{-}1)f_0/2$, were explained by the Duffing oscillator model of the membrane. This model failed to predict the harmonics above 10 kHz in *dizi* tones, which might be attributed to the response of the jet to the high-frequency harmonics generated by the membrane. The second formant at 10¡V14 kHz suggested a coupling of the first transverse pipe mode.

The timbre of *dizi* was studied in terms of its stationary spectral features. The global pattern of spectral envelopes of *dizi* tones, which is correlated to the auditory attribute ¡§brightness¡¨, was distinguished from envelope jaggedness due to a predominance of odd-numbered harmonics. In addition to harmonics, subharmonics also play a role in the perception of *dizi* tones.

The bright timbre of *dizi* tones is due to the harmonics accumulating around the two formants, whose contributions to loudness, stream segregation and breathiness were discussed. The upper notes in the second octave of the *dizi* are characterized by a poor spectral content. When a *dizi* melody ascends from bright notes to dull notes, the sound source seems to move to a remote place and become more difficult to localize. These *spatial effects* in *dizi* music were related to current theories about sound source localization.

The *dizi* tones with weak even-numbered harmonics are characterized by the hollow and/or nasal sound qualities described by Helmholtz. Nasal voices produced with a water membrane in the nasal cavity have the spectral feature of a predominance of upper odd-numbered harmonics, by which the membrane functions as a Duffing oscillator. These upper odd-numbered harmonics can induce multi-pitches. For instance, when the spectrum of a *dizi* tone is dominated by 7th, 9th, and 11th harmonics, it could produce three pitches: f_0, $9f_0/4$, and $9f_0/5$. This effect, which can enrich the texture of a *dizi* melody, was quantitatively studied using a pitch model based on autocorrelation analysis.

The *dizi* tones in the second octave sometimes contain subharmonics, which induce a rough quality. In solo music for the large *dizi* this quality is important for imitating singing. Large *dizi* tones with subharmonics often recall melancholic voices. Whereas past psychoacoustic models show difficulties in

roughness calculation for sounds with subharmonics, this thesis provided a new model for the perception of subharmonics, which takes into account the grouping mechanisms in auditory scene analysis.

The Chinese Membrane Flute (*dizi*): Physics and Perception of its Tones

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor philosophiae
(Dr. phil)

eingereicht an der Philosophischen Fakultät III
der Humboldt-Universität zu Berlin

von

Chen-Gia Tsai
geboren am 06.04.1971 in Yünlin (Taiwan)

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jürgen Mlynek

Dekanin der Philosophischen Fakultät III
Prof. Dr. Ingeborg Baldauf

Gutachter: 1. Prof. Dr. Wolfgang Auhagen
2. Dr. René Caussé

Tag der mündlichen Prüfung: 28.11.2003

Abstract

The goal of this thesis is to address the acoustic and psychological foundations of the music of the Chinese membrane flute (*dizi*). The *dizi* has a membrane covering a hole in the wall of the instrument between the mouth hole and the uppermost finger-hole. The membrane can enhance upper harmonics of *dizi* tones, but it also restricts the tone range. Reductions in the playability of notes in the third octave were explained by input impedance calculations and measurements. Wrinkles in the membrane are critical to the production of the characteristic *dizi* timbre. Theoretical and experimental studies showed that wrinkling can reduce the membrane's cubic non-linearity, thereby avoiding the jump phenomenon of a Duffing oscillator.

Another effect of the membrane is the inhomogeneity of the *dizi* timbre. As the membrane is driven by the acoustic pressure in the pipe and radiates sounds, it vibrates more when pressure antinodes lie below it. According to spectral features, the *dizi* tone range can be divided into five registers. Three major spectral features, (1) the first formant at 4–6 kHz, (2) the predominance of odd-numbered harmonics, and (3) high-frequency subharmonics $(2n\text{-}1)f_0/2$, were explained by the Duffing oscillator model of the membrane. This model failed to predict the harmonics above 10 kHz in *dizi* tones, which might be attributed to the response of the jet to the high-frequency harmonics generated by the membrane. The second formant at 10–14 kHz suggested a coupling of the first transverse pipe mode.

The timbre of *dizi* was studied in terms of its stationary spectral features. The global pattern of spectral envelopes of *dizi* tones, which is correlated to the auditory attribute "brightness", was distinguished from envelope jaggedness due to a predominance of odd-numbered harmonics. In addition to harmonics, subharmonics also play a role in the perception of *dizi* tones.

The bright timbre of *dizi* tones is due to the harmonics accumulating around the two formants, whose contributions to loudness, stream segregation and breathiness were discussed. The upper notes in the second octave of the *dizi* are characterized by a poor spectral content. When a *dizi* melody ascends from bright notes to dull notes, the sound source seems to move to a remote place and become more difficult to localize. These *spatial effects* in *dizi* music were related to current theories about sound source localization.

The *dizi* tones with weak even-numbered harmonics are characterized by the hollow and/or nasal sound qualities described by Helmholtz. Nasal voices produced with a water membrane in the nasal cavity have the spectral feature of a predominance of upper odd-numbered harmonics, by which the membrane functions as a Duffing oscillator. These upper odd-numbered harmonics can induce multi-pitches. For instance, when the spectrum of a *dizi* tone is dominated by 7th, 9th, and 11th harmonics, it could produce three pitches: f_0, $9f_0/4$, and $9f_0/5$. This effect, which can enrich the texture of a *dizi* melody, was quantitatively studied using a pitch model based on autocorrelation analysis.

The *dizi* tones in the second octave sometimes contain subharmonics, which induce a rough quality. In solo music for the large *dizi* this quality is important for imitating singing. Large *dizi* tones with subharmonics often recall melancholic voices. Whereas past psychoacoustic models show difficulties in roughness calculation for sounds with subharmonics, this thesis provided a new model for the perception of subharmonics, which takes into account the grouping mechanisms in auditory scene analysis.

Keywords:

Flue instruments, impedance, Duffing oscillator, timbre, sound source localization, predominance of odd-numbered harmonics, pitch, subharmonics, roughness, auditory scene analysis

Zusammenfassung

Die Arbeit behandelt die akustischen und psychologischen Grundlagen der Musik der chinesischen Flöte (*dizi*), die sich durch eine zwischen Anblasloch und erstem Griffloch eingefügte Membran auszeichnet. Die Membran verstärkt höhere Obertöne des *Dizi*-Klangs, obwohl sie den Tonbereich verkleinert. Spielbarkeitsreduktionen der Töne in der dritten Oktave können durch Eingangsimpedanz-Berechnungen und -Messungen erklärt werden. Die Falten in der Membran sind für die Erzeugung der charakteristischen Klangfarbe der *Dizi* wichtig. Theoretische und experimentelle Untersuchungen bestätigen, dass die Falten die kubische Nichtlinearität der Membran reduzieren können. Dadurch wird das Sprungphänomen eines Duffing-Oszillators beseitigt.

Die andere Auswirkung der Membran ist die Ungleichmäßigkeit der *Dizi*-Klangfarbe. Die Membran, die durch den akustischen Druck innerhalb des Rohrs angeregt ist, schwingt stärker in dem Fall, dass Druckbäuche unter ihr liegen. Nach den spektralen Merkmalen kann der *Dizi*-Tonbereich in fünf Register aufgeteilt werden. Die drei wichtigsten spektralen Merkmale: (1) der erste Formant, (2) die Dominanz ungeradzahliger Teiltöne, und (3) hochfrequente Subharmonische, werden durch das Membranmodell des Duffing-Oszillators erklärt. Das Modell kann allerdings nicht die beobachtbaren Teiltöne oberhalb von 10 kHz voraussagen. Diese hochfrequenten Anteile entstehen durch Strömungsvorgänge im Anblasloch in Interaktion mit der Membran. Der zweite Formant im Frequenzbereich von 10–14 kHz deutet eine Koppelung der ersten transversalen Rohr-Mode an.

Die *Dizi*-Klangfarbe wird hinsichtlich ihrer stationären spektralen Merkmale untersucht. Die globale Eigenschaft spektraler Envelopen von *Dizi*-Klängen, auf die das auditive Attribut "Helligkeit" zurückgeführt werden kann, wird von der durch Dominanz ungeradzahliger Teiltöne verursachten Hüllenzackigkeit unterschieden. Neben Harmonischen spielen Subharmonische eine Rolle in der Wahrnehmung des *Dizi*-Klangs.

Die helle Klangfarbe der *Dizi* wird von den auf zwei Formanten konzentrierten Obertönen verursacht. Deren Auswirkungen auf Lautheit, Stromtrennung (stream segregation), und "Hauchigkeit" werden untersucht. Die höheren Töne in der zweiten Oktave der *Dizi* zeichnen sich durch einen armen Spektrumgehalt aus. Falls eine *Dizi*-Melodie von hellen Tönen zu dumpfen steigt, klingt die Signalquelle entfernt und schwerer lokalisierbar. Diese räumlichen Effekte werden zu vorhandenen Theorien über die Signalquellen-Lokalisierung in Beziehung gesetzt.

Der *Dizi*-Klang mit schwachen geradzahligen Obertönen ist durch die von Helmholtz beschriebenen hohle und/oder nasale Qualitäten charakterisiert. Die nasale Stimme, die mit einer aus Wasser gebildeten Membran in der nasalen Höhle produziert wird, hat die spektrale Eigenschaft der Dominanz höherer ungeradzahliger Teiltöne, wobei die Membran als Duffing-Oszillator funktioniert. Die höheren ungeradzahligen Teiltöne können mehrfache Tonhöhen produzieren. Als Beispiel kann ein von den 7. 9. und 11. Obertönen beherrschter *Dizi*-Klang drei Tonhöhen erzeugen: f_0, $9f_0/4$, und $9f_0/5$. Der Effekt, der das Gewebe der *Dizi*-Melodie anreichern kann, wird mit einem auf der Autokorrelationsanalyse basierenden Tonhöhenmodell untersucht.

Manchmal haben die *Dizi*-Klänge in der zweiten Oktave Subharmonische, die eine rauhe Qualität hervorrufen. In der Solomusik der Bass-*Dizi* ist die Qualität wichtig zur Nachahmung des Gesangs. Die Klänge der Bass-*Dizi* mit Subharmonischen deuten melancholische Stimmungen an. Während frühe Psychoakustikmodelle Schwierigkeiten mit der Rauhigkeitsberechnung von Klängen mit Subharmonischen haben, bietet die Arbeit ein neues Modell der Wahrnehnung von Subharmonischen, das Grupppierungseffekte in der auditiven Szene-Analyse berücksichtigt.

Schlagwörter:

Labialflöte, Impedanz, Duffing-Oszillator, Klangfarbe, Signalquellen-Lokalisierung, Dominanz ungeradzahliger Teiltöne, Tonhöhe, Subharmonische, Rauhigkeit, auditive Szene-Analyse

摘要

本論文旨在探討中國笛樂的聲學及心理學基礎。中國笛的管身，在吹孔與指孔之間有一膜孔，上面貼以笛膜。此笛膜的主要作用在於加強笛音的高頻泛音，不過它也導致了音域的縮小，這可由輸入阻抗（input impedance）的計算與測量得到證實。笛膜的皺紋對於笛音影響甚鉅，本研究發現皺紋可以降低笛膜的立方非線性項，防止 Duffing 振子的跳躍現象（jump phenomena）發生。

笛膜的另一個效應是使音色變得不均勻，隨音高而變化。由於笛膜受管內低頻駐波所驅動，當其位於壓力波之波腹時，會有較大的振動。依頻譜特徵的不同，笛子的音域可以分為五個音區。笛音的三個主要頻譜特徵：位於 4–6 kHz 的第一共振峰（formant）、奇數倍泛音優勢（predominance of odd-numbered harmonics）、次泛音（subharmonics），皆可由笛膜的 Duffing 振子模型來解釋。此模型無法解釋 10 kHz 以上的泛音分布，這些泛音可能跟笛膜產生的高頻對於噴流（jet）的影響有關。笛音的第二共振峰，則顯示了笛管第一橫向共振（the first transverse pipe mode）的作用。

本論文從笛音的靜態（stationary）頻譜特徵來探討笛子的音色。頻譜的「巨觀形狀」跟「局部特徵」被分開討論，前者與明亮度（brightness）有關，後者則與相鄰奇、偶數倍泛音間的強度差有關。除了泛音之外，次泛音對於笛音音色也有重要的影響。

笛音的明亮度來自於分布於兩個共振峰附近的泛音，本論文探討它們對於音量、聲部分離（stream segregation）、氣音強度（breathiness）等的影響。由於第二個八度的最高幾個笛音具有異常微弱的高頻泛音，所以當笛樂旋律上行到此音區時，突兀的音色變化可以帶來音源「遠離」或「縹緲」的錯覺，這些空間效應（spatial effects）可以由現有的聽音辨位（sound source localization）理論來解釋。

奇數倍泛音比偶數倍泛音來得強的笛音，具有 Helmholtz 提出的空洞度（hollowness）與鼻音（nasality）這兩個音質。本論文指出，當鼻腔中有水膜時，可發出具有高頻奇數倍泛音優勢的假聲嗓音，此膜具有 Duffing 振子的作用，可產生高頻的奇數倍泛音。這些高頻奇數倍泛音可以導致多重音高（multi-pitch）的現象，例如當笛音中的第 7、9、11 泛音特別強時，它可產生三個音高：f_0、$9f_0/4$、$9f_0/5$。這個現象可用 autocorrelation 的音高模型來做定量分析。

第二個八度的笛音有時會包含頻率為 $(2n\text{-}1)f_0/2$ 的次泛音，它們引起了嘈雜度（roughness），這個音質可讓低音大笛模仿幽咽、悲傷的歌聲。本論文中引用音景分析（auditory scene analysis）中匯整機轉（grouping mechanism）的觀念，提出了一個嘈雜度的新模型，以彌補心理聲學（psychoacoustics）在解釋次泛音聽覺現象上的不足。

獻給我的父母

Contents

Chapter 1. Introduction **1**
1.1 Flue instruments 1
1.2 Membrane-flue instruments 2
1.3 Goal and Framework 2
1.4 Methodology and Scope 3
1.5 Thesis outline 5
Chapter 2. Chinese flutes and their music **9**
2.1 Introduction 9
2.2 Terminology 10
2.3 History 11
2.3.1 Membrane-less flute 11
2.3.2 Membrane flute 12
2.4 Construction 13
2.5 Fingering, temperament and key-mode 14
2.6 Southern and northern styles 16
2.6.1 Traditional repertory 16
2.6.2 Performance techniques and embellishments 17
2.7 New trends in 20th century China 18
2.7.1 Repertory 18
2.7.2 Instrument evolutions 19
2.7.3 New performance techniques 20
Chapter 3. Acoustic effects of the membrane **23**
3.1 Introduction 23
3.2 Empirical rules of membrane-sealing 25
3.3 Pitch and playability 26
3.4 Inhomogeneity of the *dizi* timbre 26
3.4.1 First register 27
3.4.2 Second register 28
3.4.3 Third register 30
3.4.4 Fourth register 32
3.4.5 Fifth register 33
3.4.6 Overview of the five registers 35
3.5 Conclusions 35
Chapter 4. Input impedance of the pipe-membrane resonator **37**

4.1 Introduction 37
4.2 Piston approximation.....38
4.2.1 Equations of motion 38
4.2.2 Input impedance of a pipe-membrane system.....41
4.2.3 Admittance reductions, resonance shifts and splitting.....43
4.2.4 Resonant frequency, mass and damping coefficient of the membrane 46
4.3 Experimental study.....47
4.3.1 Experimental setup.....47
4.3.2 Materials 48
4.3.3 Experimental procedure 49
4.4 Results and discussion.....50
4.4.1 Resonance shifts and admittance reductions 50
4.4.2 Impedance correlates of performance practice.....54
4.5 Conclusions 60
Chapter 5. Non-linearity of the membrane: Wrinkling and performance practice 61
5.1 Introduction 61
5.2 Membrane's non-linearity and wrinkles.....62
5.2.1 Unwrinkled membrane.....62
5.2.2 Wrinkled membrane.....64
5.3 Theory.....69
5.3.1 Cubic non-linearity and jump phenomenon 69
5.3.2 Harmonic balance.....70
5.3.3 Worman's theorem.....72
5.4 Experimental study.....73
5.4.1 Experimental setup.....73
5.4.2 Materials 74
5.4.3 Experimental procedure 75
5.5 Results and discussion.....76
5.5.1 Magnitudes of the lowest three harmonics.....76
5.5.2 Response curves 78
5.6 Jump phenomenon, wrinkling and performance practice 79
5.7 Conclusions 83
Chapter 6. Spectral features of *dizi* tones and the membrane's cubic non-linearity 85
6.1 Introduction 85
6.2 Spectral features of *dizi* tones 89
6.2.1 Tones radiated from the membrane and the mouth hole.....89
6.2.2 *Dizi* tones in the second and third registers 90
6.3 Methods 93
6.3.1 Experimental study.....93

6.3.2 Numerical study 96
6.4 Comparisons between numerical and experimental results 97
6.4.1 Harmonic generation.....97
6.4.2 Subharmonic generation 103
6.5 Membrane location and the *dizi* timbre 107
6.6 Aerodynamics and radiation at the mouth hole.....109
6.6.1 Drive mechanism of flue instruments.....109
6.6.2 Jet's sensitivity to high harmonics.....112
6.6.3 Transverse resonance of the *dizi* pipe.....114
6.7 Conclusions 115
Chapter 7. Perception of harmonics I: Global spectral envelope 117
7.1 Introduction 117
7.2 Frequency regions 118
7.2.1 Auditory brightness: Definition and biological implications.....118
7.2.2 Center frequencies of the two formants.....120
7.2.3 First formant: Stream segregation and loudness 120
7.2.4 Second formant: Sharpness and breathiness 121
7.3 Radiation directionality.....121
7.4 Spatial effects induced by brightness changes 123
7.4.1 Distancing effect and sound absorption of air.....124
7.4.2 Localization-confusion and monaural spectral cues 127
7.5 Brightness in *dizi* music 129
7.5.1 Depiction of sunrise 129
7.5.2 Spatial effects combined with vibrato 129
7.5.3 Rapid changes in brightness.....130
7.6 Conclusions 130
Chapter 8. Perception of harmonics II: Predominance of odd-numbered harmonics.....133
8.1 Introduction 133
8.2 Helmholtz's hollowness and nasality revisited 135
8.2.1 Hollow sounds in music.....135
8.2.2 Psychological basis of hollow sound quality 136
8.2.3 Nasality in speech and music.....137
8.2.4 Nasal voices produced by a membrane 137
8.3 Timbre perception 139
8.3.1 Odd-numbered harmonics predominance in timbre spaces.....139
8.3.2 Grating in harmonic template 140
8.3.3 Critical bandwidth correlates of hollowness/nasality.....141
8.4 Pitch perception.....143
8.4.1 Residue pitch of double-spaced components 143

8.4.2 Autocorrelation model for pitch extraction and pitch strength estimation.....144
8.4.3 Multi-pitches in *dizi* solo music.....147
8.5 Conclusions 150
Chapter 9. Perception of subharmonics: Roughness.....155
9.1 Introduction 155
9.2 Roughness: models and difficulties 156
9.2.1 Psychoacoustic models and neurophysiological evidence.....156
9.2.2 Pleasant beatings in *dizi* music 158
9.2.3 Pleasant beatings in voices.....159
9.2.4 Unpleasant consonance: Roughness induced by subharmonics.....159
9.2.5 Failure of psychoacoustic models in roughness calculation 161
9.3 Experimental study.....162
9.3.1 Method.....162
9.3.2 Results and discussion 163
9.4 Modeling roughness induced by subharmonics 165
9.4.1 Auditory scene analysis and grouping mechanisms.....165
9.4.2 Model description.....166
9.4.3 Implications for future research.....168
9.5 Subharmonics in large *dizi* music 168
9.5.1 Subharmonic generation of the large *dizi* 168
9.5.2 Perception of subharmonics in large *dizi* music.....169
9.6 Conclusions 170
Chapter 10. Conclusions.....173
10.1 Physics of the *dizi*.....173
10.1.1 Impedance correlates of *dizi* tone range 173
10.1.2 Membrane-wrinkling correlates of *dizi* timbre and playability 174
10.1.3 Non-linear mechanism correlates of spectral features of *dizi* tones 174
10.2 Perception of *dizi* tones 175
10.2.1 Brightness in *dizi* music 176
10.2.2 Auditory correlates of the predominance of odd-numbered harmonics.....176
10.2.3 Roughness induced by subharmonics.....177
10.3 Perspectives 177
10.3.1 Aerodynamics 177
10.3.2 Biomusicology 178
Appendix A. Autocorrelogram for multi-pitch transcription 179
Curriculum Vitae 186
Publications.....187
Eidestattliche Erklärung 188

Chapter 1.
Introduction

1.1 Flue instruments

Flue instruments are musical instruments with similar sounding mechanisms of flutes. They include flutes, recorders, organ flue pipes, panpipes, *shakuhachis*, ocarinas and others. In a flue instrument, the sound is produced by mutual interaction between a vibrating jet and acoustic waves in a resonator. The jet blown by the musician travels along the mouth of the resonator towards a sharp edge. When the instrument produces a tone, the jet oscillates with the period of the tone. At the sharp edge, the jet is directed alternately toward the inside and the outside of the resonator. According to a potential flow theory and a vortex sound theory, this pulsing injection induces an equivalent pressure difference across the mouth (Verge 1994, Dequand 2000, Meissner 2002). The pressure fluctuations excite the resonator and sustain periodic acoustic waves. The acoustic field at the mouth in turn drives the jet with the frequency corresponding to the excited mode of the resonator. In this feedback loop, the radiation and visco-thermal losses are compensated for by the energy supplied by the jet.

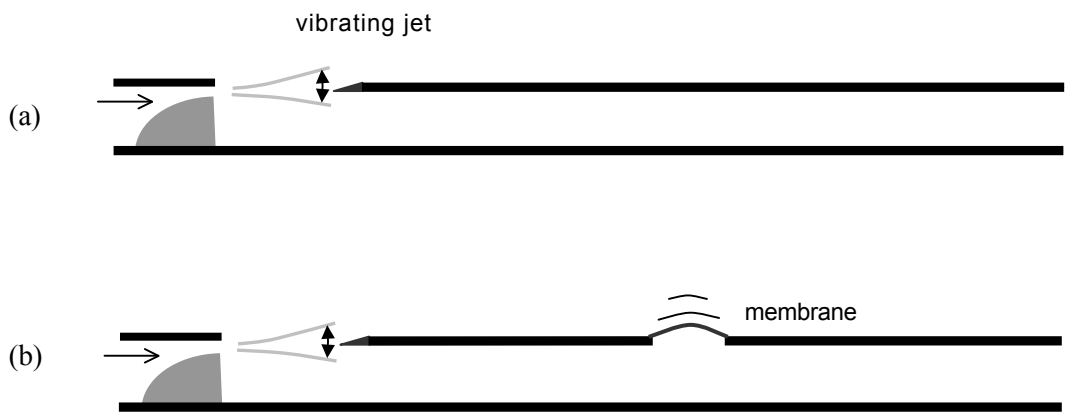


Figure 1.1: *(a) An example of flue instruments: a recorder-like organ pipe. (b) An example of membrane-flue instruments: a recorder-like organ pipe with a membrane.*

1.2 Membrane-flue instruments

Membrane-flue instruments are flue instruments with a membrane covering a hole in the wall of their resonators. The membrane is driven by the acoustic pressure in the resonator and radiates sounds when it vibrates. A comparison of flue instruments and membrane-flue instruments is illustrated in Fig. 1.1. For simplicity, a recorder-like organ pipe is chosen.

In contrast to the abundance of flue instruments in numerous musical cultures, only a few membrane-flue instruments exist. The most important of them may be the Chinese flute, the *dizi*, and the Korean flute, the *taegum*. They have a membrane hole in the wall of the instrument about halfway between the mouth hole and the uppermost finger hole. In the Hornbostel and Sachs classification system of musical instruments, the *dizi* and the *taegum* can be listed as 421.121.12, although this does not indicate the presence of the membrane. Membrane-flue instruments possess attributes of both *aerophones* and *membranophones*.

1.3 Goal and Framework

As the study of a musical instrument, this thesis is closely related to the tradition of organology (*Instrumentenkunde* in German), which was deeply rooted in historical musicology and then involved in comparative musicology after taking into account non-European musical instruments.

Organology in comparative musicology was, in essence, a descriptive science. The appearance, evolution, performance practice and musical aesthetics of an exotic musical instrument may be described and compared to related Western instruments. Its construction dimensions, intonation or sound spectra may be measured. Nevertheless, linkages between its physical properties and performance practice, between its timbre and musical aesthetics are often missing. Solid scientific bases of organology demand the methodology of systematical musicology.

The goal of this thesis is to address the acoustic and psychological foundations of *dizi* music. The framework of this thesis is illustrated in Fig. 1.2. Chapter 2 follows the methodology of organology by providing a general description of the *dizi*. Chapter 3 takes a close look at the role of the membrane in *dizi* music, posing the questions which I try to answer in the following chapters.

The physics and perception of *dizi* tones will be analyzed within the framework of systematical musicology. Chapters 4 and 5 deal with the physics of the *dizi* related to the performance practice. Chapter 6 aims to explain the spectra of *dizi* tones by modeling the membrane as a Duffing oscillator. It thus serves as transition in this thesis from the physics to the timbre perception of *dizi* tones. Chapters 7–9 aim to explore the auditory attributes and effects correlated to the spectral features of *dizi* tones, highlighting the role of the *dizi* timbre in the perception of *dizi* music.

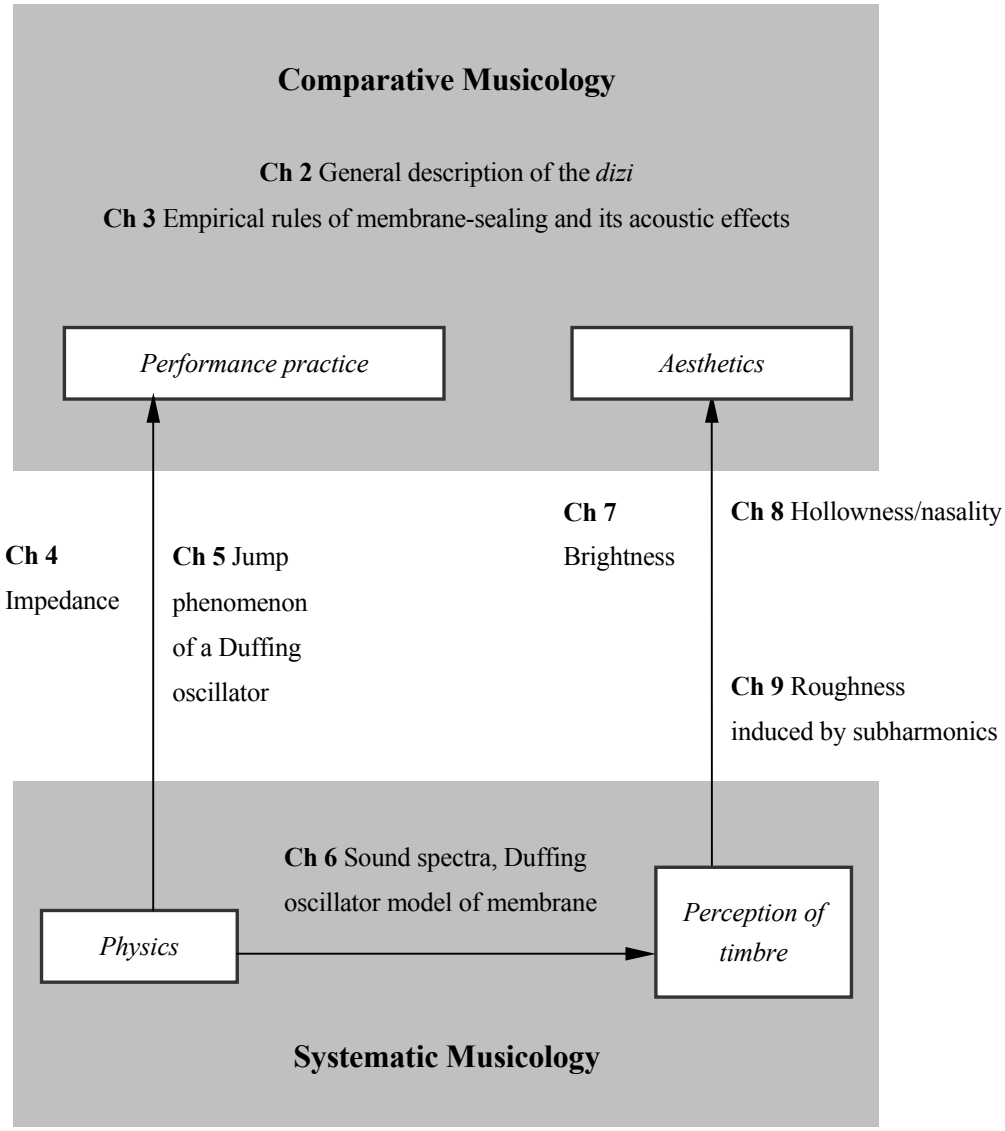


Figure 1.2: *Framework of this thesis.*

1.4 Methodology and Scope

As a consequence of the multidisciplainary nature of systematical musicology, the present study on the *dizi* is divided into two parts dealing with the physics (chapters 4–6) and perception (chapters 7–9) of *dizi* tones. In addition to the essential methods of acoustics and psychoacoustics I also intend in this thesis to connect with recent progress made in aeroacoustics, psychological acoustics and physiological acoustics.

The mechanics involved in *dizi* sound production depends on the vibration amplitude of its membrane. In small-amplitude vibrations of the membrane, the resonator of the *dizi* behaves linearly.

Chapter 4 concentrates on impedance measurements and calculations of the *dizi* resonator, which is a coupled system consisting of the pipe and the membrane. In medium-amplitude vibrations of the membrane, the resonator of the *dizi* behaves non-linearly. The membrane's non-linearity of lower orders is explored in chapter 5 by means of a harmonic balance technique. In a typical playing range, there are two non-linear mechanisms in the *dizi*: the aerodynamics at the mouth hole and the membrane in large-amplitude vibrations. In chapter 6 the membrane's vibration is simulated by a numerical method. Theories of aeroacoustics are taken into consideration in discussing the sounding mechanism and harmonic generation of the *dizi*.

While these methods mentioned above are widely available for the physics of musical instruments, the methodology in musical perception is relatively diverse and often a hotly debated issue. In this thesis, a central aspect for exploring the perception of *dizi* tones is the multidimensionality of musical timbre.

Timbre is an auditory attribute considered as a perceptual parameter of sound that is complex and multidimensional. This multidimensionality makes it impossible to measure timbre on a single continuum. Multidimensional scaling (MDS) has been a fruitful tool for studying the perceptual relationships among musical sounds and for analyzing the underlying attributes used by subjects when making (dis)similarity judgments on pairs of musical sounds. The object of MDS is to reveal relationships among a set of musical sounds by representing them in a low-dimensional Euclidean space so that the distances among the musical sounds reflect their relative dissimilarities. In MDS research, two common dimensions have been quantified psychophysically in terms of (1) transient attack and (2) spectral centroid (Grey 1977, Iverson and Krumhamsl 1993, Krimphoff *et al.* 1994, McAdam *et al.* 1995). Other dimensions such as spectral irregularity and spectral flux are shown to be less robust (McAdam and Winsberg 2000).

The present thesis concentrates on the static spectral features of *dizi* tones and does not take into account its temporal envelop, which is believed to play a crucial role in timbre recognition. Despite this scope restriction, the spectral timbre of the *dizi* itself is multidimensional and complex. Because the spectral distribution of *dizi* tones can range from 200 Hz to 22 kHz, even brightness itself may be multidimensional and should not be represented by the spectral centroid. In exploring *dizi* tones, I find the importance of some spectral features in musical perception, such as the average amplitude difference between adjacent odd/even-numbered harmonics (auditory hollowness or nasality) and distribution of subharmonics at high frequencies (auditory roughness). From this viewpoint, this thesis seems to expand the method of MDS by proposing new dimensions in the timbre space. However, it is not the main purpose of this thesis to refine MDS. It should be noted that I refer to the aspect of timbre's multidimensionality, but no experiments of (dis)similarity judgments are performed for constructing a *dizi* timbre space. The reason for this methodological adaptation lies in substantial differences between Western music and Chinese music.

Whereas MDS was developed to represent different timbres characteristic of different Western musical instruments in a timbre space, the present thesis accounts for different timbres of the same

instrument. The spectra of *dizi* tones vary a lot through the entire tone range. Inhomogeneity of the *dizi* timbre is well known by musicians. It reflects a general aesthetic tendency in traditional Chinese music that a skillful musician should produce various timbres with a solo instrument. In comparison to Western music, the score of traditional Chinese solo music is always fairly simple. During a performance, the listener concentrates on nuances in timbre. In *Qin* music, the most important musical genre of Chinese literati, an intense interest of timbre was embodied in a number of techniques dealing with various ways to pluck a string. Similarly, the varying timbre of *dizi* tones is well appreciated by the traditional music audience. As the conception of musical timbre in Chinese music differs substantially from that in Western music, the comparative approach can throw new light on musical timbre research. Specifically, the highly inconsistent spectral timbre of the *dizi* demands a rethinking of the constancy of musical timbre for instrument identification.

Inhomogeneity of the *dizi* timbre leads to *spatial effects* in *dizi* music, which can be related to current theories about sound source localization. The upper notes in the second octave of the *dizi* are characterized by a poor spectral content. When a *dizi* melody ascends from bright notes to dull notes, the sound source seems to move to a remote place and become more difficult to localize (Lin 1997). These spatial effects will be explained in terms of a theory of auditory distance perception and monaural cues of sound localization.

Another interesting phenomenon in *dizi* music is the rich 'texture' of a single *dizi* tone. Due to some spectral features, *dizi* tones sometimes evoke multi-streams in auditory scene; the *dizi* melody sounds rich in texture. This auditory illusion will be related to the notion of *auditory scene analysis* (Bregman 1990), which breaks a sound mixture into elements and unifies proximate elements into discrete objects, or 'streams'. Grouping mechanisms in auditory scene analysis are considered to be governed by some 'grouping rules'. The rich texture of a single *dizi* tone reflects a limitation of the grouping mechanism based on pitch perception. To demonstrate this effect, I follow the current models for computational auditory scene analysis to use autocorrelation analysis for pitch extraction and pitch strength estimation (e.g., Brown and Cooke 1995, Ellis 1996).

1.5 Thesis outline

The thesis begins with two introductory chapters, which are followed by the main body (chapters 4–9) dealing with the physics and perception of *dizi* tones.

In chapter 2, a general description of the *dizi* is provided. It includes the history, construction, performance techniques, repertoire and styles of the instrument. New trends of *dizi* music in 20th century China are also mentioned.

In chapter 3, the empirical rules of membrane-sealing and the acoustic effects of the *dizi* membrane are described. This chapter emphasizes the characteristics of a membrane-flute, triggering the following scientific studies.

In the first part of the main body (chapters 4–6), the physics of the *dizi* is studied.

Chapter 4 studies the linear behavior of the membrane in small-amplitude vibrations. Input impedance measurements and calculations of the *dizi* resonator are performed.

Chapter 5 explores the physics of wrinkles in the membrane and its non-linear behavior in medium-amplitude vibrations.

In chapter 6, the membrane in large-amplitude vibrations is modeled as a Duffing oscillator driven by the low-frequency acoustic waves in the pipe. This model aims to explain the spectral features of *dizi* tones in various registers.

In the second part of the main body (chapters 7–9), the perception of *dizi* tones is considered, concentrating on their steady spectral features. Chapters 7 and 8 deal with the perception of harmonics with frequencies nf_0 (where f_0 is the perceived pitch, n = 1, 2, 3...), while chapter 9 deals with the perception of subharmonics with frequencies $(2n\text{-}1)f_0/2$.

Chapter 7 studies the perception of global spectral envelop of *dizi* tones. It is represented by brightness and related to the two formants of the *dizi*.

Chapter 8 explores the perception of spectral envelope jaggedness of *dizi* tones, which is due to the predominance of odd-numbered harmonics.

In chapter 9, an auditory model for roughness induced by subharmonics is provided. *Dizi* tones and human voices containing subharmonics are compared.

References

Bregman, A.S. (1990). Auditory Scene Analysis. MIT Press.

Brown, G.J., and Cooke, M.P. (1994). Perceptual grouping of musical sounds: a computational model. Journal of New Music Research 23, 107–132.

Dequand , S. (2000). Duct Aeroacoustics: from Technological Applications to the Flute. PhD thesis, Technische Universiteit Eidhoven, The Netherlands.

Ellis, D.P.W. (1996). Prediction-driven computational auditory scene analysis. PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbre. Journal of the Acoustical Society of America 61, 1270-1277.

Iverson, P., and Krumhansl C.L. (1993). Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America 94(5), 2595-2603.

Krimphoff, J., McAdams, S., and Winsberg S. (1994). Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychophysique. Journal de Physique IV, Colloque C5, 4, 625-628.

Lin, G.F. (1997). Dikuan Yioqin.（林谷芳《諦觀有情：中國音樂裡的人文世界》，台北：望月文化）

McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research 58, 177-192.

McAdams, S., and Winsberg, S. (2000). Psychophysical quantification of individual differences in timbre perception. In: Contributions to Psychological Acoustics: Results of the Eighth Oldenburg Symposium on Psychological Acoustics (edited by A. Schick, M. Meis and C. Reckhardt), University Oldenburg, 165-181.

Meissner, M. (2002). Aerodynamically excited acoustic oscillations in cavity resonator exposed to an air jet. Acustica 88, 170-180.

Thrasher, A. (2001). "*Di*" in The New Grove Dictionary (edited by S. Sadie), Macmillan, Vol. 7, 277-279.

Verge, M.P., Caussé, R., Fabre, B., Hirschberg, A., Wijnands, A.P.J., and van Steenbergen, A. (1994). Jet oscillations and jet drive in recorder-like instruments. Acustica 2, 403-419.

Wessel, D. (1979). Timbre space as musical control structure. Computer Music Journal 3(2), 45-52.

Chapter 2.
Chinese flutes and their music

Abstract

The *dizi* is a Chinese transverse flute made from bamboo. It differs from the Western flute in having a thin membrane covering a hole in the wall of the instrument about halfway between the mouth hole and the first finger-hole. Chinese membrane-less flutes were first documented as the *chi* in the Zhou dynasty (1122–221 BC). The invention of membrane-sealing dates back to the Tang dynasty (618–907 AC), but the popularization of the Chinese membrane flute was accomplished by the rise of *Kunqu* opera in the mid-16th century.

The *qudi* and *bangdi* are two traditional types of the *dizi*. The *qudi* is the "southern" type, associated with the lyrical style influenced by *Kunqu* opera. The *bangdi*, smaller than the *qudi*, prevails in northern China with lively melodies, which require a more animated performance style with tongue techniques and glissandos.

There is no absolute standard finger-hole placement for the *dizi* and its intonation can be affected by the membrane's properties. Instead of any fixed temperament, the key-mode system was used in traditional *dizi* music. In the 20th century, this system was replaced by Western diatonic tuning system.

Several new *dizi* were invented in the 20th century. The *xindi*, a membrane-less bamboo flute, was designed to play a complete chromatic scale in Western music. The *dizi* with two membranes was designed to reduce the inhomogeneity of the *dizi* timbre. The invention of the large *dizi* has opened a new page of Chinese music, as this instrument matches the traditional Chinese literati culture particularly well. Of the new performance techniques developed in the past decades, singing into the *dizi* and playing the *dizi* with vibrating lips are relevant to the membrane.

2.1 Introduction

The purpose of this chapter is to provide a brief introduction to the Chinese flute. Following the methodology of organology, this chapter covers a number of topics such as terminology, history,

construction, performance techniques, repertory and style. As the central issue of the thesis concerns the acoustic effects of the membrane of the *dizi*, I will not go into details about every musical issue. Discussions on temperaments, performance techniques, music notation and the social/political background of Chinese flute music can be found in other papers (e.g., Zhao 1985, Rao 1987, Wang 1989, Thrasher 1978, Lau 1995). This chapter emphasizes the role of the membrane when dealing with these issues in organology.

This chapter begins with the terminology of the Chinese flute in section 2.2. Section 2.3 gives a brief survey of the history of the Chinese flute, which is divided into the era of the membrane-less flute and the era of the membrane flute. The construction of the traditional Chinese flute is described in section 2.4. Traditional Chinese flute music refers here to the era before the 20th century.

Temperament has been an overarching theme of Chinese flute research. Zhao (1985) analyzed the influences of temperature on the Chinese flute intonation. Rao (1987) compared the temperament of the Chinese flute and the ancient bells ensemble in the Zenghou Yi treasures (5th century BC). Although some interesting results have been obtained, I argue in the light of the membrane's acoustic effects that there is no standard temperament in traditional Chinese flute music. Section 2.5 shows the fingerings of the *dizi* and introduces the notion of the key-mode system, which may be more essential than the notion of fixed temperaments in traditional *dizi* music.

A list of modern music notation for the Chinese flute can be found in most textbooks for the *dizi*. However, in traditional music, written notation only offered a skeletal form of the melody for all instruments. The art of embellishments rooted in oral traditions has become an integral aspect of techniques. In section 2.6 the repertory and embellishments of the southern and northern styles of *dizi* music are described. New trends of *dizi* music in the 20th century are briefly addressed in section 2.7, which discusses instrument evolutions and new performance techniques relevant to the membrane's acoustic effects.

2.2 Terminology

The Chinese flute is often called *dizi* or *hengdi*. The term *di* is a generic term for all Chinese flutes. In the present thesis, the modern name, *dizi*, is used to refer to the Chinese membrane flute. The *hengdi*, literally 'transverse flute', is an old name dating back to the seventh century. It can refer to flutes with or without a membrane. Besides *dizi* and *hengdi*, the *zhudi* (literally 'bamboo flute') and the *zhong-guo-di* (literally 'Chinese flute') are also popular names for the Chinese flute.

The flutes accompanying various Chinese operas have different names. The *qudi* (literally 'song flute') are specifically associated with *Kunqu* opera and *Luantan* opera originating from central-eastern China. The *qudi* accompanying the *Kunqu* opera is called *kundi*. On the other hand, the *bangdi* is associated with *Bangzi* opera originating from northern China. It is also known as *duandi* (literally 'short flute') or *gaodi* (literally 'high flute'), as it is shorter than the *qudi* and its tones are higher.

The Chinese flutes can be named for their material, such as the *yudi* (literally 'jade flute'), *tiedi* (literally 'iron flute') and *zhudi* (literally 'bamboo flute'). The latter's English translation, the Chinese bamboo flute, appears frequently in musical literature. However, this name reflects the prevailing misconception that the characteristic timbre of the *dizi* is due to its material. For this reason the name 'bamboo flute' is not used in the present thesis. Others, knowing the crucial role of the membrane for the *dizi* timbre, sometimes call the instrument *modi*, literally 'membrane flute'. This name clearly indicates that the *dizi* is a membrane-flue instrument, and therefore the 'Chinese membrane flute' has been chosen for this thesis. An alternative name for the membrane flute is the "reeded flute", because the membrane is peeled from the inside of a reed. However, this name can easily be confused with reed instruments, especially the Chinese transverse reed instrument: the *bawu*.

In historical literature, the terminology of wind instruments is quite confusing. Names such as *guan* (literally 'pipe'), *chui* (literally 'blow') are often used for both flutes and reed instruments. The term *di* itself may refer to either vertical flutes or transverse flutes. Hence, historical literature should be viewed with caution.

2.3 History

2.3.1 Membrane-less flute

Flute type instruments are believed to be among the oldest melodic instruments made by humans. A flutelike fragment found in Slovenia is ascribed to the final phase of the middle Paleolithic era (50,000 to 35,000 years ago) (Kunquej and Turk 2000). The oldest Chinese flute is a bone flute from 8,000 years ago found in Wuyang County of Henan Province. It has seven holes and is still playable (Wang 1989). Another important archeological find, the Hemudu discovery, contained more than forty bone flutes from 7,000 years ago. Most of them have less than three holes, but one has six finger-holes. This bone flute closely resembles the *dizi* of today (Zhao 1985). Although these recent archeological finds have extended the history of the *dizi* back into the Neolithic era, it remains unclear whether these bone flutes have influenced the development of later Chinese flutes.

The first documented Chinese flute is the *chi* in the Zhou dynasty (1122–221 BC). Its mouth hole is located in the side of the tube either near one end or directly in the middle. In the former case, the *chi* has a similar construction to the *di*. In the latter case, the tube is closed at both ends (Chuang 1965). They were found in the Zenghou Yi treasures (5th century BC), which also contained two panpipes (Liu 1987). The *chi* is played transversely but differs from the *di* in construction, the *chi* being shorter and having a larger internal diameter. While the finger-holes and the mouth hole of the *di* lie on the same plane, the finger-holes of the *chi* lie on a surface 90 degrees from the mouth hole. Therefore, the *chi* is held with both hands on the outside of it. As the *chi* was employed in Zhou court music, it is very likely that this instrument in some way influenced the development of the *di*.

During the Han dynasty (206 BC–220 AC), the *hengchui* (literally 'transverse blow') was

imported from Central Asia. From the sixth century, transverse flutes were more commonly know as *hengdi*. Pictorial evidence and lyric resources show that this instrument was quite popular and one of the most important military instruments. Two Han *hengdi* found in the Mawangdui discovery of Hunan have six finger-holes (Wang 1989). Thrasher (1978) suggested that the Han *hengdi* may have been a flute of Indian provenance or influence. However, more research is needed before this assertion can be supported. The present-day Indian bamboo flute closely resembles the Chinese bamboo flute, but the former has no membrane.

The *hengdi*, together with the *bili* (double-reed instrument, resembling the present-day *guan*) and the *sheng* (mouth-organ), was employed in Tang (618–907 AC) court entertainment ensembles. Of the many instruments sent to Japan during this period, four *hengdi* are preserved at the *Shosoin* Repository (Hayashi 1967). These flutes have seven finger-holes and no membrane-hole. The present-day Japanese bamboo flute, *ryuteki*, is still used in the *Gagaku* orchestra. As the Japanese flute imported from China in the Tang dynasty has no membrane, it is unlikely that the Chinese membrane flute emerged or became popular as early as the Tang dynasty.

2.3.2 Membrane flute

The Chinese membrane flute was first mentioned in the early twelfth century treatise *yueshu*. The author, Chen yang, attributed the invention of membrane-sealing to a musician Liu Xi （劉係）in the Tang dynasty. The first Chinese membrane flute was named *qixing-guan*, literally 'seven-stars tube'. It became a popular instrument in the Song dynasty (960–1279 AC), while the membrane-less flutes continued to be mentioned under several different names (Thrasher 1978). It is interesting to note that a vertical membrane flute, *xiaoguan*, was also mentioned in *yueshu*. But this instrument was not documented in other literature and has never been discovered. By the following Ming dynasty (1368–1662 AC) the membrane-less flute had all but been replaced by the membrane flute. Its basic construction remains unchanged up to now.

It is noteworthy that the popularization of the membrane flute in the Ming dynasty was a consequence of the rise of *Kunqu* opera. Because the membrane brings a bright timbre to the flute, it can be clearly heard by all performers and the audience, even when the singing and the percussion instruments are fairly loud. The membrane noticeably reinforces the leading role of the *dizi* on *Kunqu* opera and the membrane-less flutes rapidly faded away. Furthermore, the rise of *Kunqu* opera in the mid-16th century highlighted the coming of the membrane flute era in Chinese music history, as many performance techniques and the musical style of the *dizi* were developed to accompany the singing in *Kunqu* opera. The lyric and embellished singing in *Kunqu* opera has deeply influenced the southern style of the *dizi*.

Before moving from this historical introduction to the Chinese membrane flute, it is important to mention another membrane flute in a musical tradition neighboring China. The Korean membrane flute, the *taegum*, is believed to be one of the three major flutes of the Unified Shilla period (668 – 935 AC). It is larger than the *qudi* and differs from Chinese membrane flutes in having a proportionally

larger membrane hole and mouth hole. As the *taegum* seems to predate the emergence of the Chinese membrane flute (Howard 1988), there is a remote possibility that the *taegum* was imported from China.

2.4 Construction

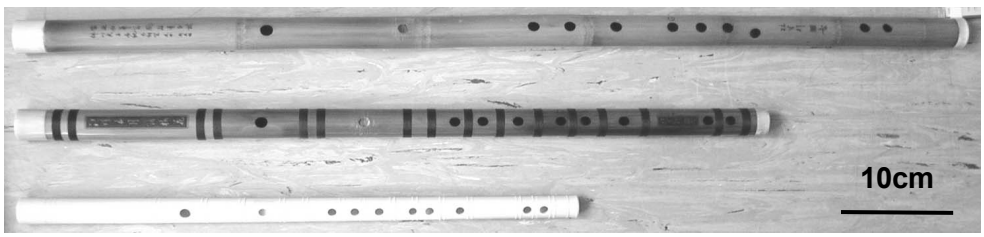
The *dizi* is a transverse flute made from bamboo. According to its size, three basic types of the *dizi* are usually differentiated. With increasing length, they are the *bangdi*, *qudi* and large *dizi*. The *bangdi* and the *qudi* are traditional instruments, while the large *dizi* was invented in the recent decades.

The *qudi* is the prevailing "southern" type, employed in musical and theatrical genres of central-eastern China. The *qudi* is externally about 40 cm or more in length. Its usual range is two octaves plus two tones – for instance, G4 to A6 for the *qudi* in C. The *bangdi* is the prevailing "northern" type, employed in musical genres of northern China. Both the *qudi* and the *bangdi* have six finger-holes, while the large *dizi* usually has seven finger-holes. These three types of *dizi* are compared in Fig. 2.1. Because the seventh finger-hole of the large *dizi* is covered by the small finger of the right hand, this hole is located a little inward.

The *dizi* has two or four closely-spaced end-holes near the foot of the instrument. These end-holes are never covered and define the effective length of the lumping air column for the lowest tone. They are sometimes used to attach one end of a cord.

The *dizi* has a thin membrane covering a hole in the wall of the instrument about halfway between the mouth hole and the first finger-hole. The membrane of the *dizi* is made of a square of thin skin peeled from the inner surface of a section of reed or bamboo. It is larger than the hole and attached with water-soluble glue.

(a)



(b)

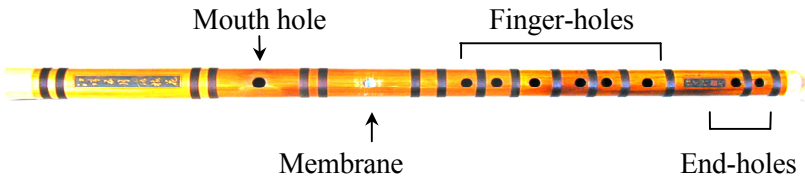


Figure 2.1: *(a) Photos of three types of the dizi: large dizi in G, qudi in C and bangdi in F. In the modern nomenclature, a dizi is named after the note played with the fingering* ●●●○○○ *or* ●●●○○○○. *(b) Construction of the dizi.*

Like the Western flute, the *dizi* has a cork sealing one end of the tube. The tube of the instrument extends about 20 cm beyond the cork, with no obvious acoustical effects, although the weight of this extension has an effect in the player's hands that is similar to that of the heavy crown plug in the head of a Western flute.

The *dizi* is normally wrapped in numerous windings of silk thread or nylon line. This helps to prevent the bamboo from splitting. The *dizi* is often tipped on both ends with bone fittings to protect the bamboo tube.

2.5 Fingering, temperament and key-mode

Table 2.1 shows the fingering table of the *dizi* given in a modern textbook for *dizi* players. Next to the absolute tone names, it also shows the solfege tone names for all fingerings in the most popular key, *Xiaogong* (●●●○○○ = do = 1). For traditional music, a sequential finger pattern of the first octave of the *dizi* tone range is basic. The second octave is fingered exactly like the first except for the only cross-fingering ○●●○○○. In the second octave it is replaced by another cross-fingering ○●○●○●. However, this tone is rarely used.

In traditional *dizi* music, there is no absolute standard in temperament or finger-hole placement, and the intervals between adjacent notes vary slightly from flute to flute. Thrasher (1978) proposed an ascending interval structure of the first octave of the *dizi*: 3/4 tone, 3/4 tone, 3/4 tone, whole tone, 3/4 tone and whole tone. He deduced this structure from measurements on four *dizi*. Three of them are traditional instruments and one is a modernized, tempered instrument. The original data of Thrasher (1978) is plotted in Fig. 2.2. It can be observed that the modernized *dizi* (marked with '∗') has been constructed closer to diatonic temperament. The ascending interval structure of the modernized *dizi* is: whole tone, whole tone, semi-tone, whole tone, whole tone, semi-tone and whole tone. On the other hand, the interval structures of the other three *dizi* are more consistent with Thrasher's proposition, namely, 150, 150, 150, 200, 150 and 200 cents.

Fingering	●●●●●●	●●●●●●	●●○●○○	●●●●○○	●●●●○○	●●○○○○	●○○○○○	○●●○○○	○○○○○○	○●●●●●	●●●●●○	●●●○○●	●●●●○○	●●●●○○	●●○○○○	●●○○○○	○○○●○○	○○○○○●	○●●●●●	●●○●○○	●○●●●●	●○○○○○
Xiaogong	5	6		7	1	2	3	4		5	6		7	1	2	3	4		5	6	7	1
Note name	G4	A4	B^b4	B4	C5	D5	E5	F5	$F^#5$	G5	A5	B^b5	B5	C6	D6	E6	F6	$F^#6$	G6	A6	B6	C7

Table 2.1: *Fingerings of the qudi in C.*

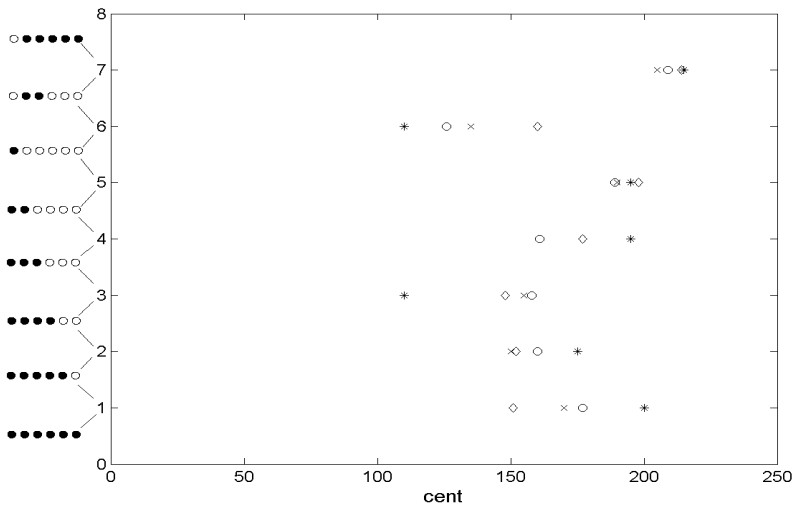


Figure 2.2: *Interval structures of the first octave measured on four dizi according to the data of Thrasher (1978). One of these dizi is modernized* (‘∗’) *and the other three are traditional. The dizi in Thrasher's collection* (‘×’) *and the dizi in the Crosby Brown Collection* (‘◇’)*, (Metropolitan Museum of Art, New York), collected on the China prior to 1894, were measured by Thrasher. The other dizi* (‘○’)*, belonging to Chuang Penli, was measured by the owner (Chuang 1968).*

Thrasher's proposition of the *dizi* temperature did not take into consideration that the pitch of tones produced by wind instruments is not completely determined by fingerings and finger-hole placement. The player's embouchure can also affect the pitch for more than 50 cent. In the case of the *dizi*, the intonation is further complicated by the membrane. A slack membrane makes most *dizi* notes lower, but a few *dizi* notes remain unaffected. Pitch shifts of *dizi* tones due to the membrane will be discussed in chapter 4. It should be noted here that pitch shifts depends on the membrane's tension and thickness. Furthermore, pitch shifts of *dizi* tones are so inhomogeneous that the first and second bore resonances for the same fingering can deviate from an octave by 50 cents or more (see Fig. 4.17).

Regarding the acoustic effects of the membrane, there is no fixed relation between the fingerings and the frequency of *dizi* tones. A good *dizi* player should modify the interval structure for various pieces. The flexibility of intonation due to the player's embouchure is intrinsic to wind instruments. Chinese musicians took advantage of this to develop the key-mode system that prevails in folk music.

The key-mode system in the *dizi*, the *suona* (a double-reed instrument) and other folk wind instruments, is a result of playing a movable solfege with the same instrument and using sequential fingerings. In the other words, it is a compromise to play various keys on a simple (key-less) instrument with sequential fingerings. Despite a number of theories on temperament and tuning systems developed in ancient China, there has been a great gap between scientific research and practice in folk music, where the limitations of instruments and the desire for a straightforward performance are more important to theoretical considerations.

Table 2.2 illustrates seven keys (‘*diao*’ in Chinese) in traditional *dizi* music. It can be seen that

each fingering can function as ‘do’ in moveable solfege. Therefore, the intervals between successive notes, such as the two tones corresponding to the fingerings ●●●●●● and ●●●●●○, cannot be invariant for playing movable solfeges. Traditional *dizi* players should change their embouchure to modify the pitch of *dizi* tones according to the key of the piece. However, it is impossible to maintain a fixed temperament for all seven keys. Consequently, each key has a quite distinct modal flavor, as correctly indicated by Thrasher (1978).

key-mode / solfege / fingerings	乙字調	正宮調	六字調	凡字調	小工調 Xiaogong	尺字調	上字調
●●●●●●	上; do	尺; re	工; mi	凡; fa	合; sol	四; la	乙; ti
●●●●●○	尺; re	工; mi	凡; fa	合; sol	四; la	乙; ti	上; do
●●●●○○	工; mi	凡; fa	合; sol	四; la	乙; ti	上; do	尺; re
●●●○○○	凡; fa	合; sol	四; la	乙; ti	上; do	尺; re	工; mi
●●○○○○	合; sol	四; la	乙; ti	上; do	尺; re	工; mi	凡; fa
●○○○○○	四; la	乙; ti	上; do	尺; re	工; mi	凡; fa	六; sol
○●●○○○	乙; ti	上; do	尺; re	工; mi	凡; fa	六; sol	五; la

Table 2.2: *Fingerings and key-mode system of the traditional dizi. The Chinese characters are the names of the key-mode. Both the traditional notation of Chinese music and the equivalent solfege are listed.*

The complicated issue of *dizi* intonation is beyond the scope of the present thesis. It should be noted that the key-mode is still used in folk *suona* music such as *Hakka Bayin* and *Beiguan* in Taiwan, whereas this tradition has faded away in *dizi* music. Nowadays, *dizi* players use a dozen *dizi* with various sizes for playing twelve keys. The modernized *dizi* is tempered according to the most popular key, *Xiaogong* (shadowed area in Table 2.2) with a diatonic scale not unlike the scale used in Western music.

2.6 Southern and northern styles

2.6.1 Traditional repertory

It is relatively straightforward to differentiate the two mainstream types of Chinese flute music: southern style and northern style, which are played by different instruments. The *qudi* is the prevailing “southern” type, employed in musical and theatrical genres of central-eastern China. The *bangdi* is the

prevailing "northern" type, employed in genres of northern China. The *qudi* is normally associated with slow and lyrical melodies originating from the singing style of *Kunqu* opera. On the contrary, *bangdi* melodies are generally lively, requiring a more animated performance style with tongue techniques and glissandos.

The *dizi* is widely used in many regional musical genres. During the Ming dynasty (1368 – 1644 AC) the *qudi* became the leading instrument in *Kunqu* opera. Nowadays, the traditional employment of the *qudi* includes *Kunqu* opera, *Jiangnan sizhu* (literally 'silk and bamboo ensemble in the lower Yangtze River region'), *Luantan* opera (originated from Shaoxing province), and Taoist rituals. The traditional employment of the *bangdi* includes *Bangzi* opera, *Errentai* (an operatic genre popular in Shanxi and Shannxi provinces, and Inner Mongolia), and *Chaozhou Xianshi* (an instrumental ensemble from the Chaozhou area in Guangdong province). Both the *qudi* and *bangdi* are used in *Gua* opera, *Hakka* opera, puppet theatre, *Beiguan* and *Bayin* (instrumental ensembles) in Taiwan.

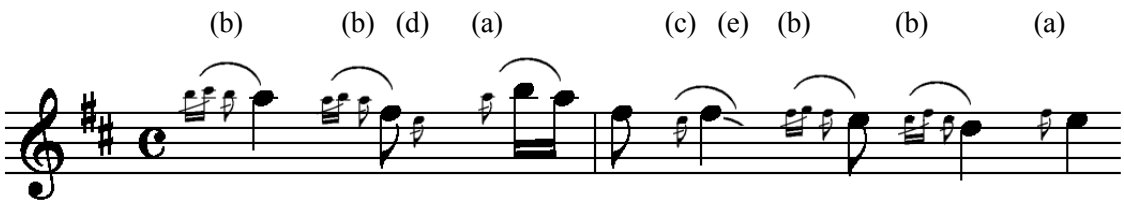
2.6.2 Performance techniques and embellishments

The *dizi* tone is produced in a similar way to the Western transverse flute. It is done by placing the lower lip just below the lower edge of the mouth hole and blowing a small stream of air across the mouth hole. Because the *dizi* mouth hole is smaller than that of the Boehm flute, the lower lip should not partially cover the mouth hole as in playing the Boehm flute, and the embouchure should be more concentrated.

There is a certain degree of improvisation in performing *dizi* music. Musicians systematically add embellishments to the 'skeletal' form of the written melody. Traditional *dizi* music genres have developed characteristic embellishments, which are idiomatic to their musical language and have become an integral aspect of *dizi* techniques. These embellishments, which vary across the southern and the northern styles, are not written in traditional notation of *dizi* music but rooted in oral traditions.

The most important embellishments used in the southern style *dizi* music are appoggiatura, trill, repeated-note and end-note decorations. They are closely related to the singing style in *Kunqu* opera.

The single appoggiatura preceding the main note can be either immediate above or below the main note [Ex.2.1(a)]. The multiple appoggiatura preceding the main note consists of a short main note and a note immediate above it [Ex.2.1(b)]. The trill is formed by the main note and the note immediate above it.



Example 2.1: *Embellishments used in the Kunqu tune 'Shanpoyang'*（山坡羊）.

In the repeated-note decoration the next higher or lower note can be used to re-articulate the same pitch [Ex.2.1(c)]. The end-note decoration is done by lifting of dropping (two or three fingers) together with a slight surge of breath. This effect can result in either a discernible end pitch [Ex. 2.1(d)] or indiscernible [Ex. 2.2(e)].

The most important embellishments used in the northern style *dizi* music are single/ double/multiple tonguing, flutter tonguing, glissando, and accented glissando. While the tongue is seldom used in the southern style *dizi* music, the single, double and multiple tonguing are basic in the northern style *dizi* music. The flutter tonguing is a method of tone production which consists in rolling the tongue. The glissando is made possible by the exposed finger-holes, which the fingers can gradually cover or uncover to change the pitch continuously. Very often musicians combine the glissando with the flutter tonguing to producing a 'fluttered glissando'. The accented glissando is done by closing all holes with fingers one after another from the sixth finger-hole to the first finger-hole rapidly and strongly, while a short and fast stream is blown into the mouth hole. Thereby the accent of the first note of the glissando is clearly heard (Wang 1989).

2.7 New trends in 20th century China

2.7.1 Repertory

Nowadays, the *dizi* repertory dominating concerts in China, Hongkong and Taiwan is different from traditional repertories. It consists almost exclusively of arrangements of existing pieces and original compositions written in recent decades. The majority of these new concerto-like compositions for the *dizi* call for dazzling finger techniques. Frequently, a modern Chinese orchestra supports *dizi* melodies with harmony that did not exist in traditional Chinese music. From a historical view, the rise of the new *dizi* repertory reflected the tide of the "National music" (in Chinese, 'guo-yue' or 'minyue') after 1940s, which was strongly influenced by Western music.

Lau (1995) related this development of *dizi* music to the process of modern nation-state formation. He divided the history after 1949 of the *dizi* repertory into three periods according to the dominant stylistic norm found in most compositions.

> In the first period (1949-1964), most of the solo *dizi* pieces are characterized by the use of the theme and variation procedure, and the music is extracted almost exclusively from various regional instrumental ensemble repertories. [...] In the second stage of its developments (1965-1978; the period generally referred to as the Cultural Revolution) many *dizi* composers [...] turned to revolutionary songs as the basis for the compositions. [...] In the third phase of *dizi* music development (from 1979 to 1989), after the fall of the Gang of Four and in the spirit of cultural liberalism [...] *dizi* composers began to incorporate new musical ideas, innovative performing techniques, and experimental formal principles in their compositions.

Lau (1995) also classified *dizi* music into the categories traditional repertory, composed repertory, and imported repertory. According to his statistics on approximately 400 *dizi* pieces, over 60% were composed repertory, while traditional and imported repertory each occupied roughly 15%.

2.7.2 Instrument evolutions

In addition to the repertory, the new trends of *dizi* music also stimulated instrument evolutions. As musical ideals shifted and new compositions called for equal-tempered scales, makers repositioned finger-holes accordingly. In the 1920s, a fully chromatic *dizi*, *xindi* (literally 'new flute') or *lüdi* (literally 'tempered flute'), was invented, with eleven finger-holes but usually without the membrane hole (Gao 1959). This membrane-less bamboo flute is especially suitable for a large, modern orchestra, where bright *dizi* tones due to the membrane are not favored because they cannot blend with other instruments. While the modern Chinese orchestra was established after the model of the Western orchestra, the *xindi* can be regarded as a symbol of the westernization of Chinese music. This instrument was designed to play all notes appearing in Western music notation and the membrane was replaced by a rigid wall to produce mellow flute tones comparable to those of the Western flute.

A more important development of the *dizi* is the invention of the large *dizi*. Large *dizi* with various sizes larger than the *qudi* have been invented in the past several decades. Among them the large *dizi* in G (●●●○○○○ = G4) is most frequently used, whose tone range is D4 to G6. Although the large *dizi* is comparable in tone range to the large membrane flute in Korea, the *taegum*, they employ different techniques and show different styles. Lin (1997) indicated that the large *dizi* matches the traditional Chinese literati culture very well. As some compositions for the large *dizi* show a philosophical impression similar to *Qin* music, this instrument has opened a new page in the music history of the Chinese membrane flute, as will be further discussed in section 9.5.

The paralleled-*di* ('peidi' in Chinese) was invented in the 1950s by Zhao Songting, a great *dizi* musician, composer and educator. He bundled two, three or four *dizi* with different sizes together, so that it is easier to switch from one *dizi* to another (Zhao 1985).

A large, curved *dizi* is used by the musician Hu Jiexu. The curved tube enables the musician to hold the long instrument with widely separated finger-holes in a more comfortable position. Large *dizi* with a U-tube has been invented after the model of the Western U-tube flute, but is rarely used.

The *dizi* with two membrane-holes was invented by the contemporary musician Li Chen. It is designed to reduce the inhomogeneity of the *dizi* timbre. Because of the enhanced acoustic effects of the two membranes, this *dizi* sounds rather buzzing, sometimes rough.

Among these recent construction developments of the *dizi*, three are relevant to the membrane. The *xindi* discards the membrane, thus losing the characteristic timbre of the *dizi*, whereas the buzzing timbre of the *dizi* is exaggerated by the *dizi* with two membranes. The rise of the large *dizi* is associated with interesting acoustic effects of the membrane on a large flute.

2.7.3 New performance techniques

Not only have the instrument themselves evolved, new performance techniques have also been developed over the past fifty years. Some of them were triggered by the performance techniques of the *suona*, a Chinese double-reed instrument with a flare. For instance, Zhao Songting introduced the suona technique of circular breathing to the *dizi*, probably based on his early experience of playing the *dizi* and the *suona* in regional music ensembles.

Playing the *dizi* with vibrating lips, just like playing a brass instrument, is similar to the advanced *suona* technique: *kachiang*. It is difficult to apply this technique to woodwind instruments, because there is no mouthpiece supporting the player's lips. Playing the *dizi* with vibrating lips produces a timbre similar to the Chinese transverse reed instrument: the *bawu*. The reedy sound results from the sympathetic vibrations of the non-linear membrane.

Two new performance techniques combine singing with the *dizi*. Singing while playing the *dizi*, and singing into the *dizi* are two different techniques. The former produces a two-voice texture consisting of the singing voice and the *dizi* tone, while the latter introduces an interesting timbre by exploiting the non-linear property of the *dizi* membrane. When singing into the *dizi*, the player's mouth entirely covers the mouth hole of the *dizi*. The singing voices are much brighter than normal human voices, recalling tones produced by the mirliton.

Two innovative techniques appear in 'Langye Shenyun' (A Wonderful Melody of Mountain *Langye*) composed and performed by Yu Xunfa. The first – whistling into the *dizi* – often produces two pitches, one corresponding to whistling and the other corresponding to the turbulence noises filtered by the pipe. The second technique in 'Langye Shenyun' is hitting the finger-holes with the fingers but without blowing. Because the air column excited by the hitting fingers has definite resonances, a pitch sensation can be evoked, although it is always rather short and weak because of fast decay. This technique is also employed in contemporary Western flute music, but it may be more suitable for flutes without keys, because hitting violently can lead to damage of the keys.

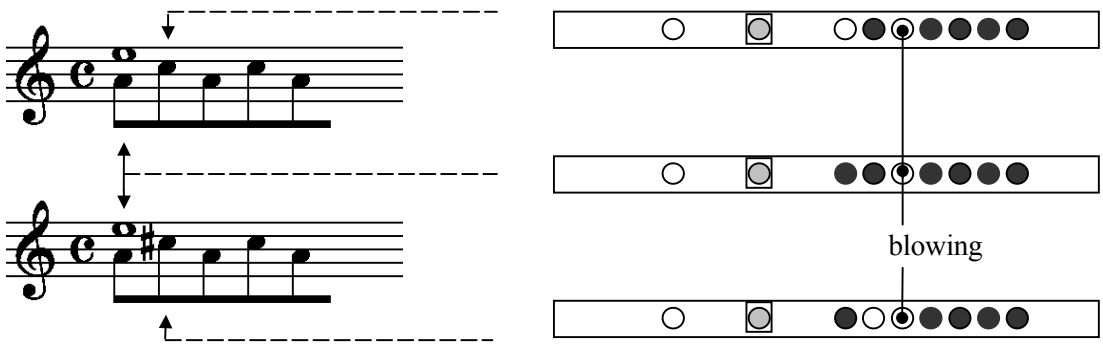


Figure 2.3: *Triad-like multi-phonics produced by blowing the third finger-hole of the dizi.*

Multi-phonics of wind instruments, sometimes used in contemporary Western music, also appear in *dizi* music. In 'Qiohu Yueyie' (A Night With Moon at Autumn Lake) composed and performed by Yu Xunfa, multi-phonics are used to imitate a multi-phonic bell sound. By blowing into the fourth finger-hole instead of the mouth hole, Jan Yuoming produces triad-like multi-phonics in his composition 'Ting Quen' (Listening to The Spring). Fig. 2.3 illustrates the notes and the fingerings for the triad-like multi-phonics.

Among these new techniques, blowing into a finger-hole cannot be applied to the Western flute with keys. Singing into the flute and playing the flute with vibrating lips are techniques closely related to the non-linear property of the membrane. Although they can be applied to the Western flute, the effects will be less remarkable because few harmonics will be generated.

References

Jan, Y.M. (1991). *Dizi Jichu Jiaucheng*.（詹永明《笛子基礎教程》，達欣出版社）

Kunej, D., and Ivan, T. (2000). New perspectives on the beginnings of music: archeological and misicological analysis of a middle Paleolithic bone "flute". In: The Origins of Music (edited by N.L. Wallin, B. Merker and S. Brown), MIT Press, 235-268.

Lau, F. (1995/1996). Individuality and political discourse in solo *dizi* compositions. Asian Music 27(1), 133-151.

Lau, F. (1996). Forever red: the invention of solo *dizi* music in post-1949 China. British Journal of Ethnomusicology 5, 113-131.

Rao, Y., Carterette, E.C., and Wu, Y. (1987). A comparison of the musical scales of the ancient Chinese bronze bell ensemble and the modern bamboo flute. Perception and Psychophysics 61(6), 547-562.

Thrasher, A. (1978). The transverse flute in traditional Chinese music. Asian Music 5, 92-114.

Thrasher, A. (2001). "*Di*" in The New Grove Dictionary (edited by S. Sadie), Macmillan, Vol. 7, 277-279.

Wang, T.C. (1989). Notes on special techniques of *dizi* (flute) performance. Chinese Music 12(3), 51-55.

Zhao, S.T. (1985). *Diyi Chunqio*.（趙松庭《笛藝春秋》，浙江人民出版社）

Chapter 3.
Acoustic effects of the membrane

Abstract

The acoustic properties of a flute are noticeably changed by replacing a small region of the rigid wall of its pipe with a membrane. The membrane makes most flute tones lower and reduces the playability of notes in the third octave, thus restricting the *dizi* tone range to two octaves plus two notes. An appropriately slack membrane is important for the *qudi* and the large *dizi* to produce characteristic buzzing low tones. But if the membrane is too slack, the *dizi* will become uncontrollable: some notes in the first and second modes are difficult to play and tend to jump to the higher play modes. To avoid this, musicians have to stretch the membrane carefully to produce a lot of lateral wrinkles in it.

Although the membrane brings a bright timbre to the flute, the *dizi* timbre is not homogenous. Because the membrane vibrates more when pressure antinodes lie below it than it does for pressure nodes, its acoustic effects vary with pitch. According to spectral characteristics, the *dizi* tone range is divided into five registers. In the first register, rich harmonics produce a buzzing sound quality. The second register is characterized by weak even-numbered harmonics. In the third register, in addition to the predominance of odd-numbered harmonics, subharmonics with frequencies $(2n\text{-}1)f_0/2$ may appear, resulting in a rough quality. *Dizi* tones in these lowest three registers always have two formants (typically centered at 4–6 kHz and 10–14 kHz). The fourth register is characterized by the absence of upper harmonics, whereas the fifth register is characterized by their dramatic recovery.

3.1 Introduction

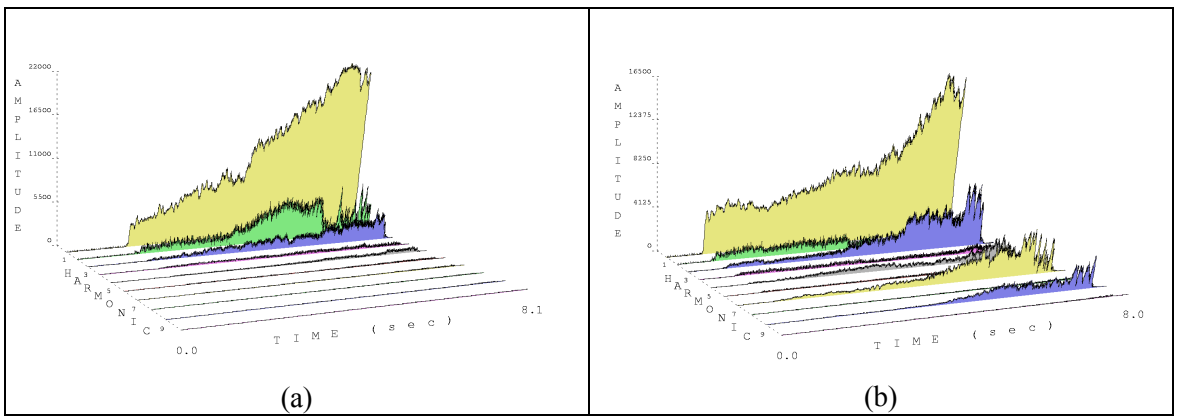
In chapter 2 a general introduction to the Chinese flute was given with the emphasis on the role of the membrane in *dizi* music. This chapter takes a closer look at the membrane's acoustic effects. The description of these effects motivates the physical and psychoacoustic studies on the *dizi* in chapters 4–9.

Although the *dizi* is famous for its bright timbre differing significantly from Western flutes, a

common misconception about the *dizi* has been that the characteristic *dizi* timbre is due to its material: bamboo. While the influence of the material on the flute timbre still remains a matter of debate, it is easy to demonstrate that the membrane is more essential than the material in producing the characteristic timbre of the *dizi*. By mounting a membrane on a finger hole of a recorder or a baroque flute, their tones will become much brighter. On the contrary, if the *dizi* membrane is completely damped with the palm of a hand, its tones will lose all harmonics above 4 kHz. The characteristic bright sound of membrane flutes owes much to the sympathetic vibration of the membrane.

As a first example, Fig. 3.1 compares the spectra of two tones produced by a *dizi* with the membrane replaced by a rigid wall surface, and with a vibrating membrane. The *dizi* was played twice with the same fingering. First, the membrane was covered with a palm, so that the membrane did not vibrate. As shown in Fig. 3.1(a), high-frequency harmonics remained fairly weak while the amplitude of the fundamental increased. In the second tone the *dizi* was played with the membrane vibrating freely. Fig. 3.1(b) shows a lot of high-frequency harmonics, whose amplitudes increased with the fundamental. The frequency of the 7th harmonics is about 5 kHz.

The purpose of this chapter is to systematically describe the membrane's acoustic effects. Since the *dizi* timbre and playability are very sensitive to the membrane's properties, section 3.2 begins by summarizing the empirical rules of membrane-sealing. The pitch shifts and playability reductions caused by the membrane are described in section 3.3. The membrane results in an inhomogeneity of the *dizi* timbre, which leads to a subdivision of its tone range into registers. In section 3.4, five registers of the *dizi* are defined according to spectral features of *dizi* tones.



(a) (b)

Figure 3.1: *Amplitude envelopes of harmonics of two crescendo tones produced by the same qudi in D with fingering* ●○○○○○. *(a) Tone produced without membrane vibration. The fundamental frequency increased from 733 Hz to 751 Hz as the intensity increased. (b) Tone produced with membrane vibration. The fundamental frequency increased from 725 Hz to 735 Hz as the intensity increased. These spectral plots were generated by J. Beauchamp's spectral analysis and display package "Sndan".*

3.2 Empirical rules of membrane-sealing

As the membrane is crucial for *dizi* sound production, one important lesson when beginning to learn to play the *dizi* is the 'art' of choosing and sealing the membrane. Although the empirical rules of membrane-sealing are well-mentioned in *dizi* textbooks, they cannot be learned without a teacher's demonstration and the student's own trail-and-error.

The *dizi* membrane is made of a square of skin peeled from the inside of reed or bamboo. It is larger than the membrane hole, and is affixed with water-soluble glue. A membrane sealed on the *dizi* can work for several months. If the tension in the membrane becomes too low, a player can increase it by applying saliva and gently spreading the edges of the paper. To get the "correct" timbre of the *dizi*, it is important to adjust the membrane with appropriate lateral wrinkles. These thin wrinkles should be distributed uniformly across the membrane and not overlap each other (see Fig. 3.2).

There is no standard method of membrane-sealing for the *dizi*. The tension and thickness of membrane desired depend on the instrument and repertoire. For *bangdi* music, commonly containing a lot of high notes, the membrane is on average thicker and sealed more tautly than that for *qudi* music. If the membrane is sealed slackly, high notes will be difficult to play. However, a taut membrane will reduce the spectral content of *dizi* tones because the amplitude or the membrane's vibration is small. In most works for the *qudi* and the large *dizi*, the membrane should be sealed more slackly, because there are few high notes but more low, soft notes. With a taut membrane, soft *dizi* tones will be too dull and the lowest two notes will lose the characteristic buzzing sound quality. The size of the membrane also affects the amplitude of the membrane's vibration. As more vibration of the membrane is desired in works for the *qudi* and the large *dizi*, their membrane holes are slightly larger than that of the *bangdi*.

Figure 3.2: *Photo of the dizi membrane.*

3.3 Pitch and playability

An immediate acoustic effect of the membrane is that the pitches of most *dizi* tones are lowered. It can be noted from Fig. 3.1 that the two *dizi* tones were played with the same instrument and fingering, but they had different pitches. In this specific experiment, the vibrating membrane made the fundamental frequency of the *dizi* tone decrease from 740 Hz to 730 Hz.

The second acoustic effect of the membrane is the playability reduction of high notes. A membrane flute has a more limited tone range in comparison to a membrane-less flute. With a vibrating membrane, the *dizi* tone range is about two octaves plus two notes. Notes beyond this range demand a very concentrated embouchure and a taut membrane, thus seldom used in traditional repertoires. If the membrane is replaced with a rigid wall, the *dizi* tone range could extend to three octaves. The membrane reduces the playability of *dizi* tones in the third octave.

Next to the membrane tension, wrinkles in the membrane are of great importance for *dizi* playability and timbre. If the membrane is sealed tautly and is free from wrinkles, the *dizi* tones will become dull as the membrane hardly vibrates. A slack membrane without wrinkles can generate bright tones, but makes the instrument uncontrollable. The *dizi* can be so unstable that it either emits no sound when being blown, or suddenly it sounds raucously. Soft *dizi* tones become impossible to play. Furthermore, the upper notes in the first play mode tend to jump to the second play mode; the lower notes in the second play mode tend to jump to the third play mode. In order to avoid these problems, musicians must stretch the membrane carefully to produce a lot of lateral wrinkles in it. Wrinkles in the membrane serve largely to stabilize the sound production of the *dizi*.

3.4 Inhomogeneity of the *dizi* timbre

For a membrane flute, the membrane is driven by the acoustic pressure in the resonator and radiates sounds when it vibrates. Because the membrane vibrates more when pressure antinodes lie below it than it does for pressure nodes, its acoustic effects vary with pitch and hence the timbre in the entire tone range is not homogenous. This inhomogeneity of timbre is inevitable for all membrane flutes, including the Korean membrane flute, the *taegum*. It is interesting to note that *taegum* players favor this inhomogeneity of timbre, considering that it makes the *taegum* a perfect solo instrument (Howard 1988). On the contrary, most of the modern *dizi* players dislike the timbre inhomogeneity of the *dizi*. Both *dizi* musicians and manufacturers have been making great efforts to remove it.

In this thesis, I divide the *dizi* tone range into five registers, which are shown in Table 3.1. These five registers are defined according to spectral characteristics of *dizi* tones. It is important to note that this division is applicable mainly to the *dizi* whose membrane is located at the midpoint between the mouth hole and the uppermost finger hole. If the membrane hole is located higher or lower, the division of register will be different.

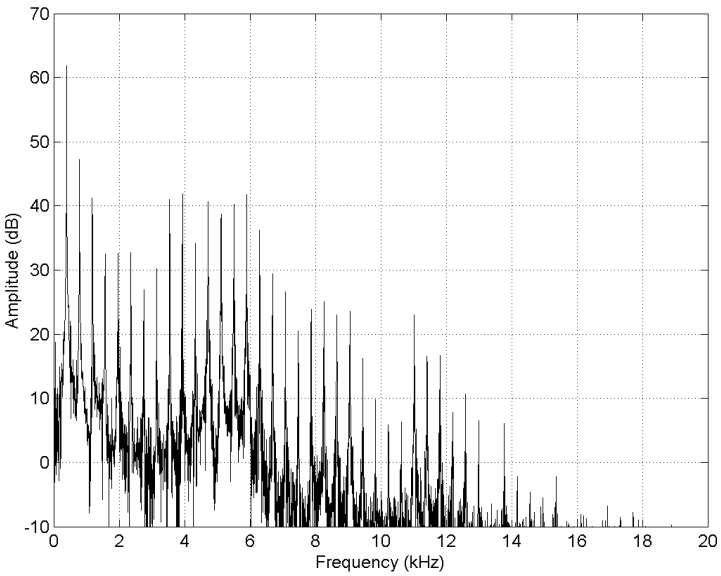
Play mode	Fingering	Note	Register
First mode	●●●●●●	G4	First register
	●●●●●○	A4	
	●●●●○○	B4	
	●●●○○○	C5	
	●●○○○○	D5	Second register
	●○○○○○	E5	
	○●●○○○	F5	
	○○○○○○	F#5	
Second mode	○●●●●●	G5	Third register
	●●●●●○	A5	
	●●●●○○	B5	
	●●●○○○	C6	
	●●○○○○	D6	Fourth register
	●○○○○○	E6	
	○●○●○●	F6	
	○○○○○○	F#6	
Fourth mode	○●●●●●	G6	Fifth register
	●●○●●○	A6	
	●○●●●●	B6	
	●○●○○○	C7	

Table 3.1: *Five registers of the qudi in C. The highest two notes B6 and C7 are difficult to play unless the membrane is very taut. They are seldom used in traditional repertoires.*

3.4.1 First register

The *dizi* tones in the first register have a smooth spectral envelope, which shows a formant typically centered between 4 kHz and 6 kHz. This formant was first mentioned by Horner *et al.* (1999).

Fig. 3.3 shows a spectrum of the lowest tone in the first register. A second formant centered at 12 kHz can be observed. If the tone intensity is higher, the harmonics above 10 kHz will be stronger.



(a)

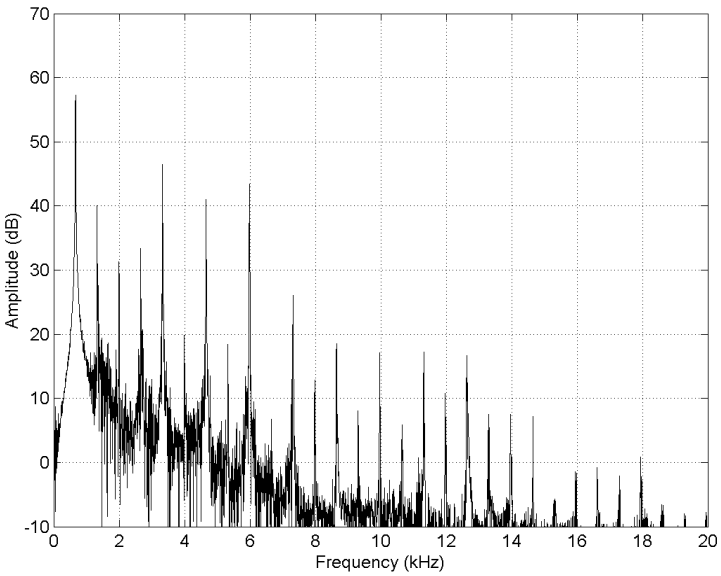
(b)

Figure 3.3: *Spectrum of a soft note in the first register of the qudi and the location of its membrane. (a) Spectrum of G4 produced by a qudi in C. First play mode, fingering* ●●●●●●*. Two formants at 5 kHz and 12 kHz can be observed. (b) Location of the membrane in relation to the fundamental acoustic pressure wave.*

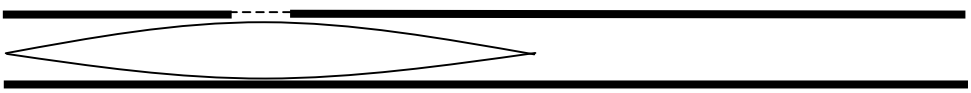
3.4.2 Second register

The *dizi* tones in the second register also have two formants, one typically centered at 4–6 kHz and the other at 10–14 kHz. More important, the odd-numbered harmonics above 2 kHz are significantly stronger than the flanking even-numbered harmonics, especially around the two formants. Fig. 3.4(a) shows a "jagged" up-and-down spectral envelope due to weak even-numbered harmonics.

The large amplitude differences between the odd/even-numbered harmonics brings a sound quality that can recall the panpipes or the clarinet, whose tones are characterized by the predominance of the odd-numbered harmonics. The relevance of Helmholtz's auditory hollowness and nasality to this sound quality will be dealt with below (see chapter 8).



(a)



(b)

Figure 3.4: *Spectrum of a note in the second register of the qudi and the location of its membrane. (a) Spectrum of E5 produce by a qudi in C. First play mode, fingering* ●○○○○○*. Two formants at 5 kHz and 12 kHz can be observed. Above 3 kHz the odd-numbered harmonics are always stronger than the flanking even-numbered harmonics. (b) Location of the membrane in relation to the fundamental acoustic pressure wave.*

It is important to note that the predominance of the odd-numbered harmonics can be reduced or eliminated by modifying the embouchure. When the jet is blown downward, the odd-numbered harmonics will no longer dominate, as shown in Fig. 3.5. This is likely to occur in *fortissimo* tones.

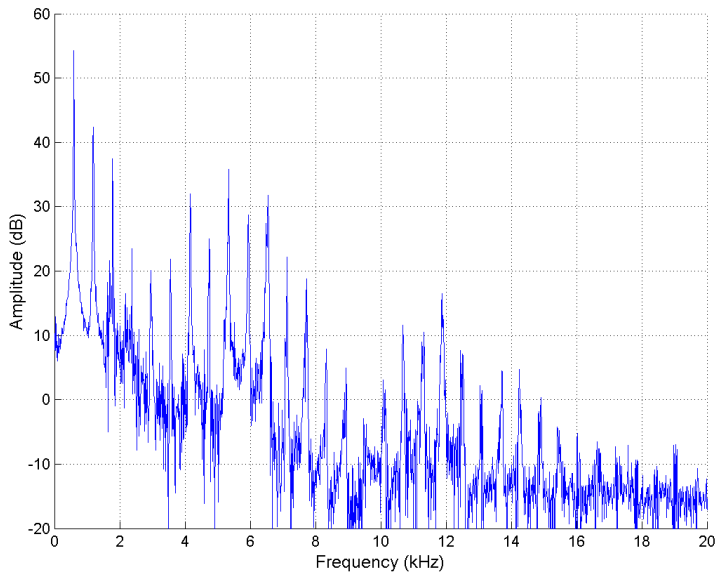


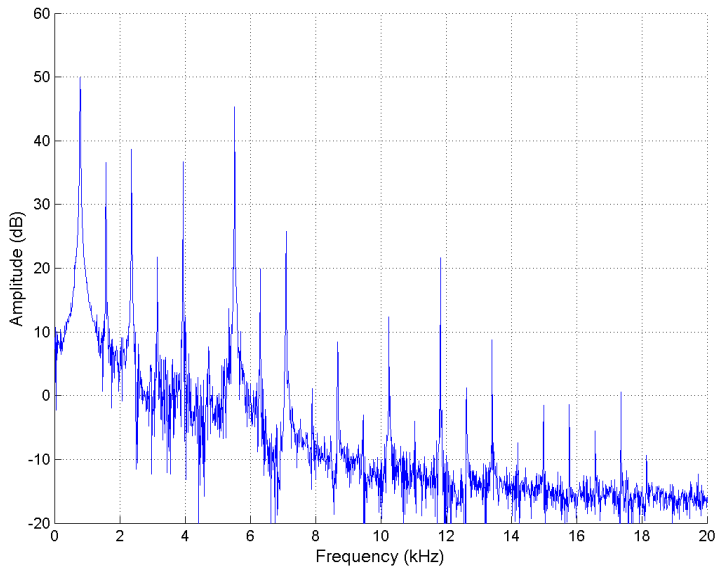
Figure 3.5: *Spectrum of E5 produced by a qudi in C (second register). First play mode, fingering* ●○○○○○. *As the jet is blown downward, there is no significant predominance of the odd-numbered harmonics.*

3.4.3 Third register

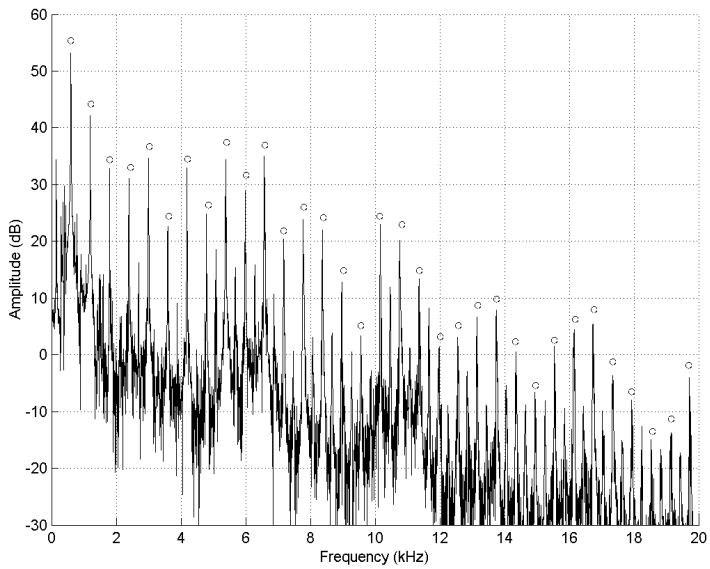
The spectra of the *dizi* tones in the third register always have two formants and are dominated by odd-numbered harmonics, like those in the second registers. These two spectral features can be observed in Fig. 3.6(a), which shows the spectrum of the lowest note in the third register produced by a *qudi* in C.

A third spectral feature, the high-frequency distribution of subharmonics, can emerge in this register with an appropriate embouchure. Fig. 3.6(b) shows the spectrum of the lowest note in the third register produced by a large *dizi* in G. It can be observed that above 4 kHz subharmonic components appear at the midpoint of pairs of adjacent harmonic components. These subharmonics, with frequencies $(2n\text{-}1)f_0/2$, characterize most soft *dizi* tones in the third register. Regarding perception of *dizi* tones, subharmonics do not make the pitch an octave lower, but bring a rough sound quality to the tones in this register, which recalls some pathological human voices. Auditory roughness induced by subharmonics at high frequencies will be discussed in chapter 9.

(a)



(b)



(c)

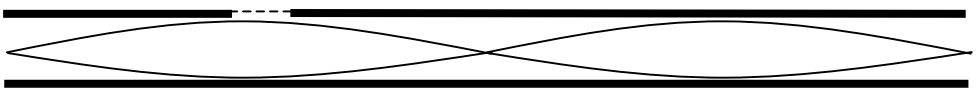
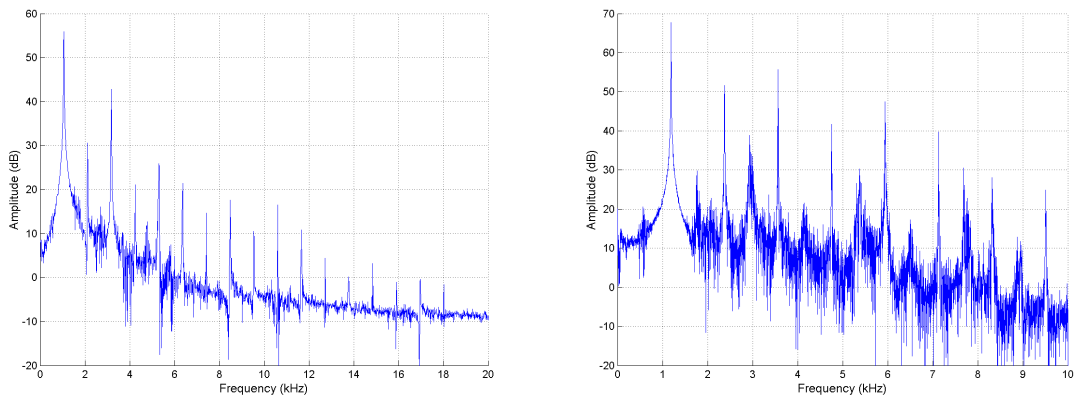


Figure 3.6: *Spectra of the lowest note in the third register of the dizi and location of its membrane. (a) G5 produced by a qudi in C. Second play mode, fingering ○●●●●●. (b) D5 produced by a large dizi in G. Second play mode, fingering ○●●●●●●. Above 2 kHz, some subharmonic spectral lines appear between harmonic spectral lines (marked with circles). (c) Location of the membrane in relation to the fundamental acoustic pressure wave.*

For the two highest notes in the third register, the predominance of odd-numbered harmonics tends to be less significant, as can be seen in Fig. 3.7(a). The spectrum of this tone does not show the two formants. Hence, it can be regarded as a transition to the next register, which is characterized by poor spectral content.

Another characteristic of the two highest notes in the third register is the widening of spectral lines of the subharmonics. This is likely to occur with a slacker membrane. Fig. 3.7(b) shows the spectrum of such a tone produced by a *qudi* in D. The subharmonic components differ from the harmonic components in spectral-line-widening. Musicians describe such an annoying sound quality as "sandy".



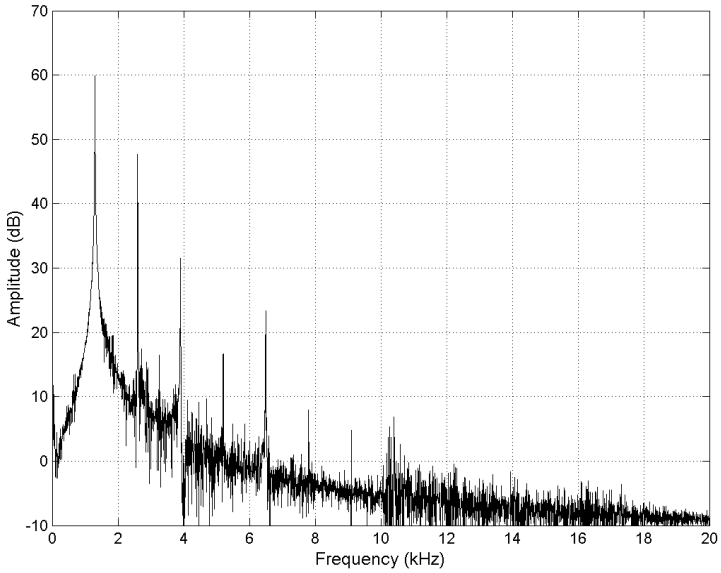
(a) (b)

Figure 3.7: *Spectra of the highest note in the third register of the dizi. (a) C6 produced by a qudi in C. Second play mode, fingering* ●●●○○○. *Significant amplitude differences between odd- and even-numbered harmonics are restricted to the range below 5 kHz. The two formants are absent as the tone intensity is low. (b) D6 produced by a qudi in D. Second play mode, fingering* ●●●○○○. *The spectral lines of the subharmonics are widened.*

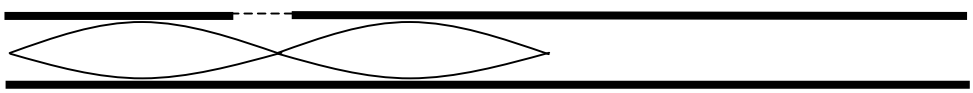
3.4.4 Fourth register

The fourth register of the *dizi* is distinct from other registers in that it has a poor spectral content. Fig. 3.8(a) shows the spectrum of such a dull tone produced by a *qudi* in C. This spectrum is dominated by the lowest two harmonics and shows no formant. Instead of harmonics, turbulence noises are distributed above 10 kHz. Dull *dizi* tones in this register thus always sound breathier than tones in other registers.

The poor spectral content can be explained by the acoustic pressure in the pipe. Fig. 3.8(b) shows the membrane location in relation to the fundamental acoustic pressure wave. In this register, the membrane is located near a pressure node of the fundamental wave. Consequently, it is driven by a weak acoustic pressure and hardly vibrates, hence not generating rich harmonics. A conspicuous absence of brightness is the most remarkable feature of the timbre inhomogeneity of the *dizi*.



(a)



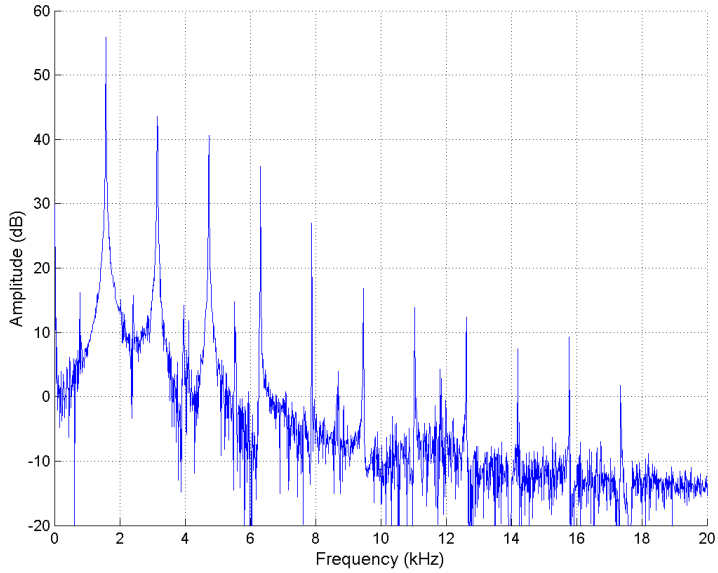
(b)

Figure 3.8: *Spectra of a note in the fourth register of the dizi and location of its membrane. (a) E6 produced by a qudi in C. Second play mode, fingering* ●○○○○○*. The spectrum shows no formant and the harmonics decay rapidly as its frequency increases. (b) Location of the membrane in relation to the fundamental acoustic pressure wave.*

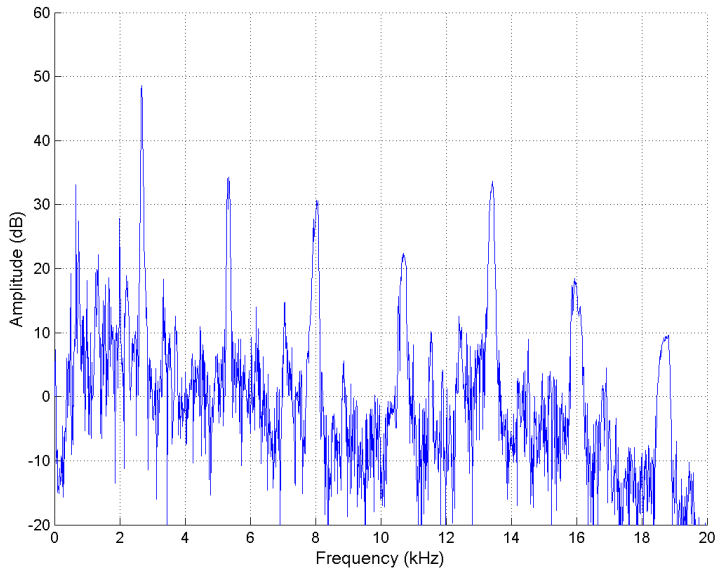
3.4.5 Fifth register

The fifth register is characterized by a striking recovery of brightness. The lower tones of the fourth play mode have rich harmonics, because the membrane is located near a pressure antinode of the fundamental. Fig. 3.9(a) shows the spectrum of a tone in the fifth register produced by a *qudi*. The frequency range of the harmonics distribution extends to 18 kHz. For the *bangdi*, whose tone range is higher than the *qudi*, the tones in the fifth register sound even more powerful. Fig. 3.9(b) shows the spectrum of a short *forte* tone produced by a *bangdi*. The harmonics above 12 kHz are fairly strong.

(a)



(b)



(c)

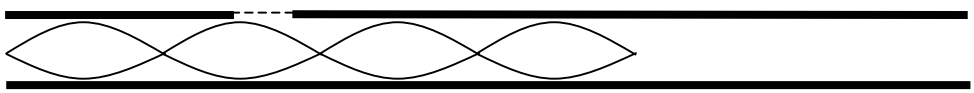


Figure 3.9: *Spectra of notes in the fifth register of the dizi and location of its membrane. (a) G6 produced by a qudi in C. Fourth play mode, fingering ○●●●●●. (b) E7 produced by a bangdi in A. Fourth play mode, fingering ○●●●●●. Spectral components below 2 kHz are produced by other instruments. (c) Location of the membrane in relation to the fundamental acoustic pressure wave.*

3.4.6 Overview of the five registers

Some concluding remarks should be made regarding the division of registers. The five registers of the *dizi* are divided according to three spectral features: (1) the distribution of harmonics at high frequencies, (2) the average amplitude difference between flanking odd/even-numbered harmonics, and (3) the distribution of subharmonics at high frequencies. Fig. 3.10 illustrates the variation of these three features across the five registers. These curves are drawn according to informal estimation. It can be noted that the transitions from the first to the second and from the third to the fourth register are gradual, whereas the transitions from the second to the third and from the fourth to the fifth register are sharp.

It is important to note that both the predominance of odd-numbered harmonics and the subharmonics distribution depend on the embouchure. Consequently, these two spectral features are not as robust as the distribution of harmonics at high frequencies.

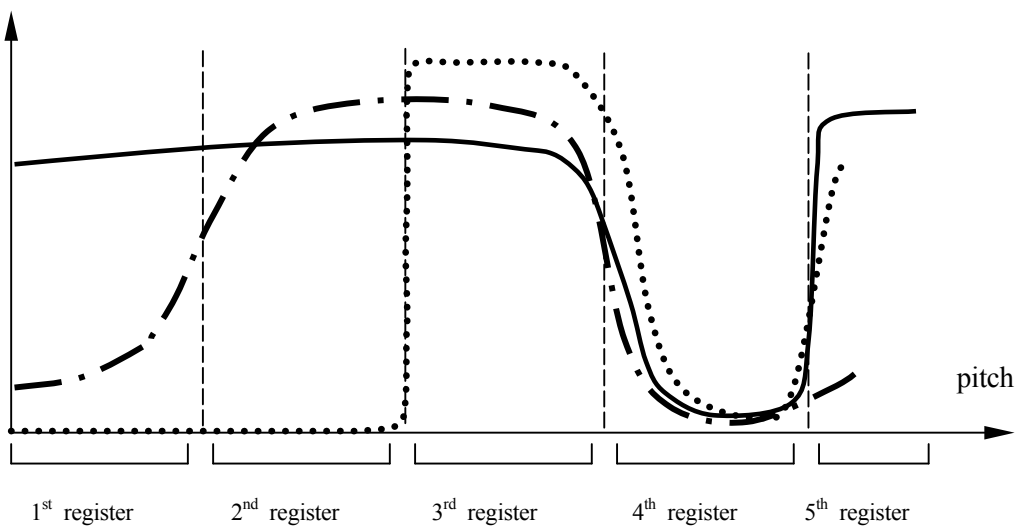


Figure 3.10:. *Three spectral features varying across the five registers of the dizi: distribution of upper harmonics (—), average amplitude difference between adjacent odd/even-numbered harmonics (—·), and distribution of upper subharmonics (···).*

3.5 Conclusions

The acoustic effects of the *dizi* membrane include (1) pitch reductions (2) tone range restriction, and (3) timbre changes. The vibrating membrane of the flute will make most pitches lower and reduce the playability of the notes in the third octave. To play *bangdi* music containing high notes, the membrane should be thicker and sealed more tautly than that for a *qudi*. A slack membrane is important for the *qudi* to produce buzzing low tones. However, if the membrane is too slack, the *dizi*

will be uncontrollable. The upper notes in the first play mode tend to jump to the second play mode, while the lower notes in the second play mode tend to jump to the third play mode. Furthermore, soft tones are impossible to play. Lateral wrinkles in the membrane largely stabilize the sound production of the *dizi*.

The membrane results in *dizi* timbre inhomogeneity. According to spectral characteristics, the *dizi* tone range is divided into five registers. In the first register, dense harmonics extending beyond 10 kHz produce a bright, buzzing sound quality. The second register is always characterized by weak even-numbered harmonics. The predominance of the odd-numbered harmonics can be reduced by modifying the embouchure. In the third register, subharmonics with frequencies $(2n\text{-}1)f_0/2$ can appear. While the *dizi* tones in the lowest three registers always have two formants centered at 4–6 kHz and 10 -14 kHz, the fourth register is characterized by the absence of rich harmonics. The fifth register is characterized by a recovery of brightness and the harmonics distributed above 10 kHz can be so strong that the *dizi* tones are fairly sharp.

References

Horner, A., Ayers, L., and Law, D. (1999). Synthesis modeling of the Chinese *dizi*, *bawu*, and *sheng*. Journal of the Audio Engineering Society 47(12), 1076-1087.

Howard, K. (1988). Korean Musical Instruments: a Practical Guide. Se-Kwang Music Publishing.

Lin, G.F. (1997). Dikuan Yioqin.（林谷芳《諦觀有情：中國音樂裡的人文世界》，台北：望月文化）

Thrasher, A. (2001). "*Di*" in The New Grove Dictionary (edited by S. Sadie), Macmillan, Vol. 7, 277-279.

Chapter 4.
Input impedance of the pipe-membrane resonator[1]

Abstract

The linear behavior of the *dizi* resonator, a pipe-membrane system, is studied by modeling its membrane as a piston. Input impedance calculation of a pipe-membrane system shows some phenomena common in soundboard-string systems. Resonance shifts and admittance reductions predicted by this linear piston model are consistent with the impedance measurements of a simple pipe-membrane system. Impedance measurements of the *qudi* in C show resonance shifts caused by the membrane of up to 60 cents. Musicians can compensate for the intonation distortions due to the membrane by modifying the embouchure. Typically, the resonant frequency of the membrane f_m lies between 2 kHz and 3 kHz for the *qudi* and the large *dizi*. Experimental results show that admittances of the bore resonances of the *qudi* in C supporting the notes B6 and C7 are largely reduced by the membrane, because the vibration modes of the pipe-membrane resonator of the *dizi* are damped by the membrane when the pipe's resonances are near f_m. Consequently, the *dizi* range is restricted to about two octaves plus two notes. Since f_m sets an upper limit of the tone range, the membrane of the *bangdi* tends to be sealed more tautly than that of the *qudi* for producing high notes.

4.1 Introduction

Since Bernoulli, the flute has been known to oscillate with the fundamental frequency corresponding to one of the resonances of the pipe. These resonances are minima of its input impedance curve (i.e., maxima of the admittance curve). This contrasts with reed or lip-reed instruments, whose resonances are maxima of the impedance curve.

For decades, it has been standard to measure or to calculate the impedance curves of woodwind

---

[1] This chapter is based on work performed at the Acoustics Laboratory of the University of New South Wales, with the assistance of Joe Wolfe and John Smith.

instruments to estimate pitches corresponding to various fingerings. This approach is also basic for the present study of the *dizi*, as no such measurement has been published. More important, the impedance curve depends not only on the fingering, but also on the state of the membrane. Input impedance measurements in various states of the membrane are expected to reveal the influence of the membrane on the linear behavior of the *dizi*.

The purpose of this chapter is to explain the effects of the membrane on pitch and playability by impedance calculation and measurements. As mentioned in chapter 3, the membrane makes the pitches of most *dizi* notes lower and the playability of the tones in the fifth register decreases with the tension of the membrane. A slack and thin membrane brings a bright timbre to low notes at the expense of tone range.

When the wavelengths of *dizi* tones are much longer than the size of the membrane, only the (01) mode of the membrane is excited. In this chapter, the membrane is modeled as a piston attached to a linear spring. The equations of motion of the *dizi* resonator are derived and its impedance is calculated in section 4.2. Section 4.3 addresses experimental studies on impedance measurements of a simple pipe-membrane system and the *dizi*. The results are related to *dizi* performance practice in section 4.4.

4.2 Piston approximation

4.2.1 Equations of motion

The radius of the membrane hole of the *dizi* is much smaller than the wavelengths of the tones produced by the instrument; $a/\lambda = o(10^{-1})$. Over the range of frequencies studied, it is expected that only the (01) mode of the membrane is excited. In this chapter, the *dizi* membrane is modeled as a piston sealing the membrane hole and mounted on a spring. This piston model is illustrated in Fig. 4.1.

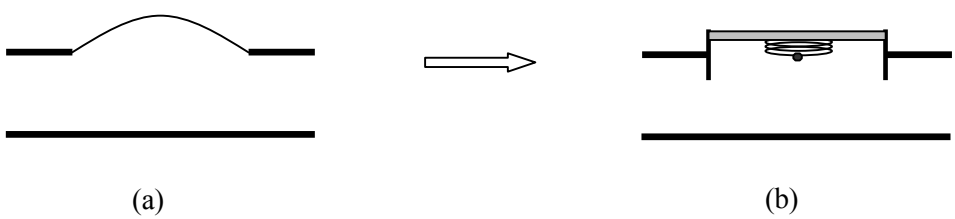


Figure 4.1: *Modeling the membrane as a piston. (a) A vibrating membrane on a hole in the wall of a pipe. (b) A vibrating piston attached to a spring.*

In a low-frequency approximation, the air loading as an end correction term is included in piston's mass and the radiation loss is ignored. The non-linearity of the piston's spring is not considered here. It is involved in *dizi* sound production when the vibration amplitude of the membrane is large. Medium- and large-amplitude vibrations of the membrane will be explored in chapters 5 and 6.

First, we use this model to derive the wave equations of the air column in the pipe, which is coupled to a piston in its wall. When the Helmholtz number based on the pipe diameter D is small, namely $He = f D/c = o(10^{-1})$, the acoustic wave in the pipe can be described in terms of plane waves. The wave equations of the acoustic pressure $p(x,t)$ and the acoustic flow $U(x,t)$ are

$$p_t = -\frac{\rho_0 c^2}{S_0} U_x \tag{4.1}$$

$$p_x = -\frac{\rho_0}{S_0} U_t \tag{4.2}$$

where $S_0 = \pi D^2/4$ is the cross-sectional area of the pipe, c the speed of sound and ρ_0 the average air density.

Eq. (4.1) describes the compressive elasticity of the air. In the presence of a piston in the wall of the pipe, this equation is modified as the moving piston introduces extra airflow. We focus on the air under the piston, which is illustrated in Fig. 4.2 as a window.

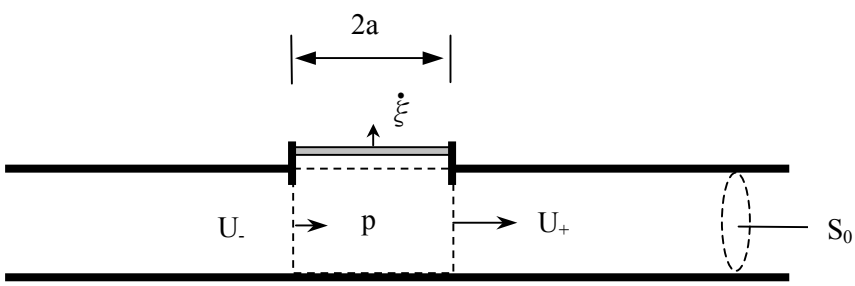


Figure 4.2: *Acoustic flow in the window under the piston. U_- and U_+ are the flows through the left side and right side of this window; ξ is the displacement of the membrane and p is the acoustic pressure in the window.*

At low frequencies, the sound propagation in this window can be neglected. Hence, the pressure in the window is assumed to be uniform; Eq. (4.2) becomes $p_x = 0$. This assumption is available only for the lower harmonics of *dizi* tones. A brief discussion on the limitation of this assumption is given in the following *intermezzo*.

Intermezzo

The assumption of a uniform pressure in the window shown in Fig. 4.2 is available only when a << λ and D << λ. At high frequencies, the vibrating membrane can induce local air density fluctuations so that the acoustic pressure in the window is no longer uniform. For example, when the Helmholtz number He = 1/4, a higher mode of the air column in the pipe can be excited. Pressure fluctuations reach its maximum at the bottom of the window, as illustrated in Fig. 4.5(a).

Fig. 4.5(b) illustrates an analogue for this caricature. Imaging a pressure source above the membrane hole and removing the membrane, we get a series of pressure wave antinodes that establish equilateral triangles. The wavelength of this mode is equal to 4D, where the membrane serves as an acoustic flow source instead of an acoustic pressure source. For a large dizi in G, D=16mm, the eigenfrequency of this mode is approximately 5 kHz. Regarding high-frequency harmonics possibly produced by the membrane, the assumption of a uniform pressure in the window is only available in discussing the fundamental and several low-order harmonics of dizi tones. Higher modes of the air column in the pipe will be further discussed in section 6.5.

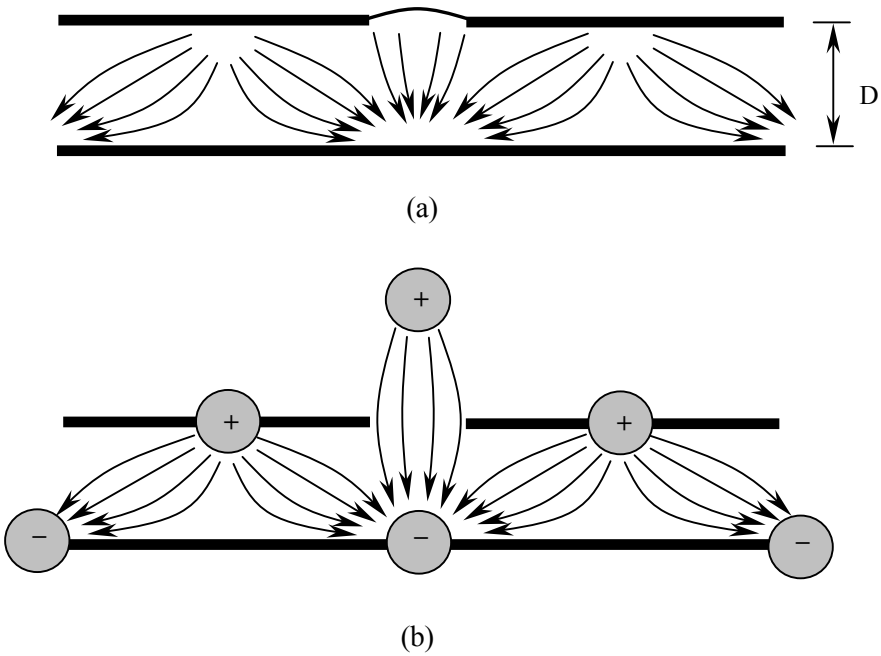


Figure 4.3: *A transverse mode of the pipe excited by the vibrating membrane in its wall. Helmholtz number He = fD/c = 1/4. The arrows represent acoustic flow streamlines. (a) The membrane as a flow source. (b) In an analogue the membrane is replaced by a pressure source above the membrane hole. The circles represent pressure antinodes and their phase relations are labeled by plus and minus symbols.*

From Fig. 4.2, the total flow out the window is

$$U_{out} = U_{+} - U_{-} + S_m \dot{y} \approx 2aU_x + S_m \dot{x} \tag{4.3}$$

where $S_m = \pi a^2$ is the area of the circular piston. For the air in the window, Eq. (4.1) becomes

$$p_t = -\frac{\rho_0 c^2}{S_0}(U_x + \frac{\pi a}{2}\dot{x}) \tag{4.4}$$

Eq. (4.4) represents a source term of acoustic pressure waves in the pipe corresponding to the piston's velocity. It is important to note that the vibrating piston induces an acoustic flow difference at the left and right sides of the window, thus serving as a pressure source. The effective pressure in the window induced by the piston is in proportion to its displacement y.

$$p_{eff} = -\frac{\rho c^2}{S_0}\frac{\pi a}{2}x \tag{4.5}$$

Finally, the motion of the piston must be described by Newton's equation,

$$m\ddot{x} + R\dot{x} + m\omega_m^2 x = S_m p \tag{4.6}$$

where m is the piston's mass, R its damping coefficient, and ω_m its resonant angular frequency. Eqs. (4.4) and (4.6) are the motion equations of the piston and the air under it at low frequencies. It can be noted from Eq. (4.6) that when the piston is located at a pressure node of an acoustic wave, it hardly vibrates.

4.2.2 Input impedance of a pipe-membrane system

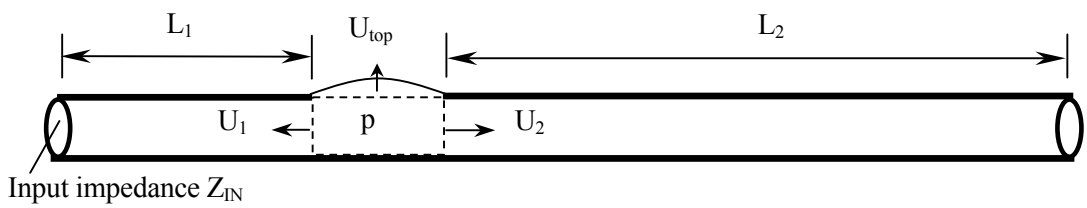


Figure 4.4: *A pipe with a membrane sealed on a hole in its wall.*

The input impedance of a pipe-membrane system depends on the location of the membrane. Consider the pipe-membrane system illustrated in Fig. 4.4.

Referring to the impedance of an open pipe, we obtain

$$U_2 = \frac{p}{jZ_0} \tan kL_2 \tag{4.7}$$

Referring to Eq. (4.6), the mechanical impedance of the membrane modeled as a mass-spring system is

$$Z_{mechanical} = \frac{F}{\dot{x}} = R + \frac{jm(w^2 - w_m{}^2)}{w} \tag{4.8}$$

The membrane's acoustic impedance is related to its mechanical impedance by the formula

$$Z_{acoustic} = \frac{p}{U_{top}} = \frac{Z_{mechanical}}{S_m{}^2} \tag{4.9}$$

At low frequencies, the sound propagation in the window can be neglected. So $U_2 = - U_1 - U_{top}$. Since

$$U_{top} = p / Z_{acoustic} \tag{4.10}$$

the impedance at the left side of the window is

$$Z_2 = \frac{p}{U_2} = \frac{-1}{\dfrac{1}{jZ \tan kL_2} + \dfrac{1}{Z_{acoustic}}} \tag{4.11}$$

The input impedance of a pipe with a general termination can be calculated when the impedance of its termination is known. The input impedance Z_{IN} of the pipe with length L_1 is related to Z_2 by Eq. (8.23) in [Fletcher & Rossing 1991].

$$Z_{IN} = Z_0 \left[\frac{Z_2 \cos kL_1 + jZ_0 \sin kL_1}{jZ_2 \sin kL_1 + Z_0 \cos kL_1} \right] \tag{4.12}$$

The expression of Z_{IN} in terms of L_1, L_2 and $Z_{acoustic}$ can be obtained by substituting Eqs. (4.8), (4.9) and (4.11) to Eq. (4.12). The general form of Z_{IN} is complicated, but if the membrane is exactly located at the middle of the pipe, namely $L_1 = L_2 = L$, the input impedance becomes

$$Z_{IN} = \frac{-Z_0^{\ 2} \sin^2 kL + jZ_0 Z_{acoustic} \sin 2kL}{jZ_0 \sin kL \cos kL + Z_{acoustic} \cos 2kL} \tag{4.13}$$

It is clear that Eq. (4.13) gives the input impedance of an open pipe with length $2L$ as $Z_{acoustic} \rightarrow \infty$, while it gives the input impedance of an open pipe with length L as $Z_{acoustic} \rightarrow 0$. The former case corresponds to replacing the membrane with a rigid wall; the latter case corresponds to replacing the membrane with an ideal open hole, which is not physically realizable.

The pipes used in our experiments have radius $r_0 < 8$mm. It is thus reasonable to ignore the radiation loss but not the wall loss. The wall loss is taken into account by introducing an imaginary part of the wave number k.

$$k = \frac{w}{v_w} - j a_w \tag{4.14}$$

where ν_w is the phase velocity and α_w is the attenuation coefficient per unit length. Their values are given in Eqs. (8.14) and (8.15) in [Fletcher & Rossing 1991].

$$v_w = c\left[1 - \frac{1.65 \times 10^{-3}}{r_0 f^{1/2}}\right] \tag{4.15}$$

$$a_w = \frac{3 \times 10^{-5} f^{1/2}}{r_0} \tag{4.16}$$

4.2.3 Admittance reductions, resonance shifts and splitting

Combining Eqs. (4.13) with (4.14), we can predict the input impedance of the *dizi*. Fig. 4.5 gives a theoretical input admittance curve $1/|Z_{IN}|$ of a pipe-membrane system, whose configuration mimics the resonator of a large *dizi* in G with all finger holes open. The bore resonances are represented as peaks in this curve.

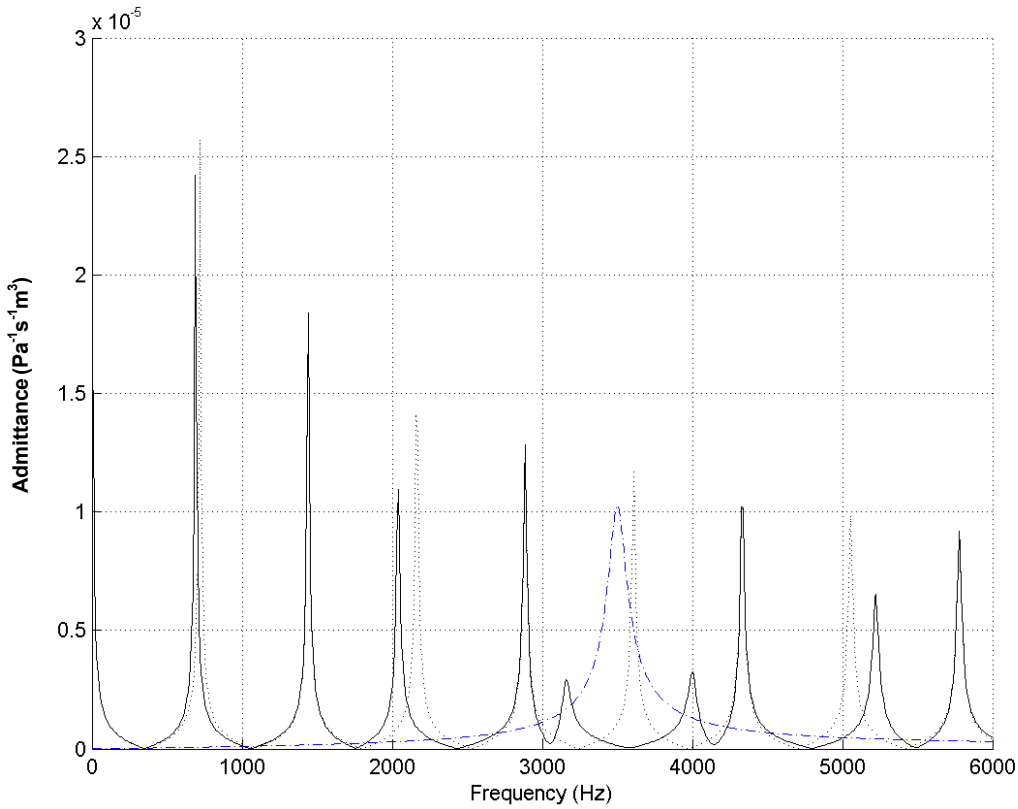


Figure 4.5: *A theoretical admittance curve of a pipe with a membrane in the wall at the midpoint. The curves for the open pipe with (—) and without (- - -) membrane vibration are compared with the acoustic admittance curve of the membrane (— ·). Resonant frequency of the membrane* $f_m = \omega_m/2\pi$ *= 3.5 kHz. Mass of the membrane* $m = 8\times10^{-7}$ *kg. Damping coefficient* $R = 6\times10^{-4}$ sec^{-1}*kg. Length of the pipe* $2L = 240$ *mm. Radius of the pipe* $r_0 = 9$ *mm. Radius of the membrane hole* $a = 5$ *mm. This configuration mimics a large dizi with the fingering* ○○○○○○○.

It can be observed that the membrane does not affect the even-numbered bore resonances of the pipe, as it is located at a pressure node of these standing waves. On the other hand, admittances of the odd-numbered resonances are reduced by the membrane, as it is located at a pressure antinode of these standing waves. The linear piston model of the *dizi* membrane predicts significant admittance reductions of the bore resonances close to the membrane's resonant frequency f_m. These odd-numbered vibration modes are damped by the membrane, whose damping is much larger than the wall loss in the pipe.

Next to the admittance reductions, the odd-numbered bore resonances are shifted away from the resonant frequency of the membrane. The fifth resonance, lying near the membrane's resonant frequency f_m = 3.5 kHz, is split into two resonances in the presence of the membrane. The lower resonance corresponds to that the membrane moves in phase with the acoustic flow wave, while the higher resonance corresponds to that the membrane moves in opposite phase with the acoustic flow

wave. Fig. 4.6 depicts the eigenfunctions of the two resonances calculated by Eq. (4.4). For simplicity, the resonance splitting here is assumed to occur at the first mode.

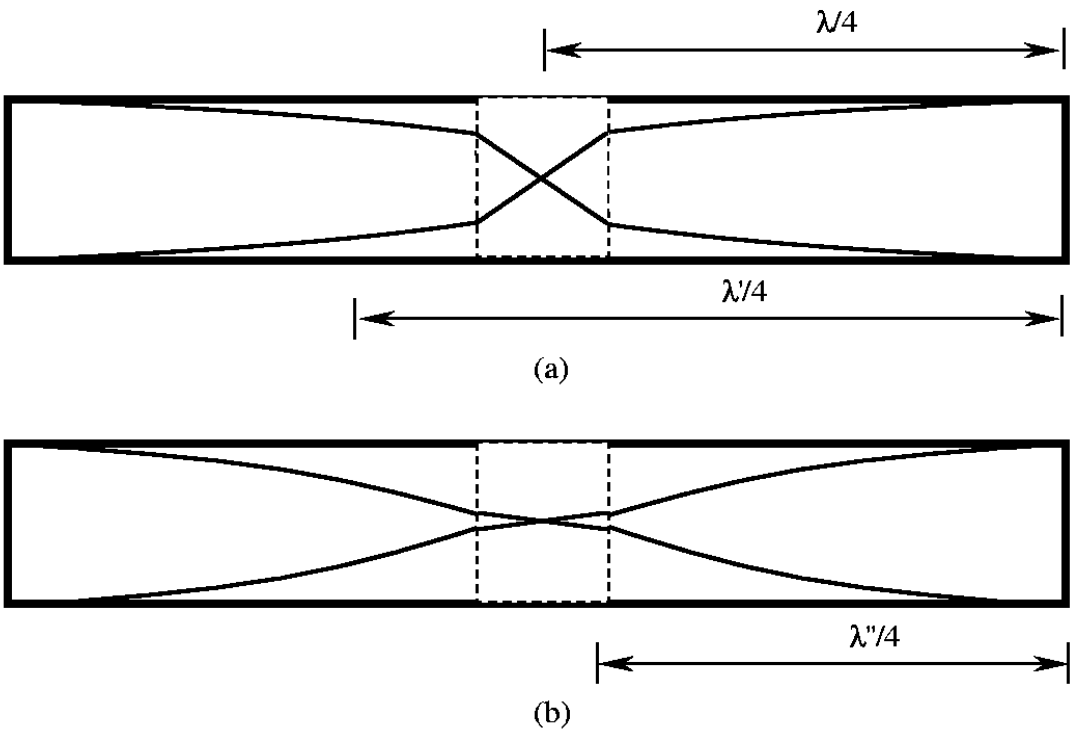


Figure 4.6: *Eigenfunctions of the air column in an open pipe with a membrane in the wall at the midpoint when the lowest resonance of the air column lies near f_m. (a) The membrane moves in phase with the acoustic flow wave. The wavelength becomes longer; $\lambda'>\lambda$. (b) The membrane moves in opposite phase with the acoustic flow wave. The wavelength becomes shorter; $\lambda''<\lambda$.*

As can be seen in Fig. 4.6, the standing waves in this pipe-membrane system are characterized by the discontinuities of U_x at the left and right sides of the window. The magnitude of U_x becomes larger or smaller as the membrane moves in phase or in opposite phase with the acoustic flow wave. These discontinuities of U_x are compensated for by the vibrating piston, which induces a airflow difference between the left and right sides of the window. In the real three-dimensional world, however, these discontinuities are smoothed.

Admittance reductions, resonance shifts and resonance splitting in a pipe-membrane system are analogous to those in the coupled system consisting of a vibrating string terminated on a soundboard. The impedance of string-soundboard systems was calculated by Gough (1981). In the weak coupling limit, the coupling to the soundboard increases the damping of string resonances, with little effect on their frequencies. In the strong coupling limit, the damping is greater. Furthermore, the frequencies of the string resonances are moved away from those of the soundboard resonance. Below the soundboard resonance, string and soundboard move in phase (a mass-like termination of the string). At higher frequency, they move in opposite phase (a spring-like termination). Resonance splitting at the

soundboard resonance gives rise to the well known wolf tone (Raman 1916, Schelleng 1963, Güth 1978).

The occurrence of the wolf tone in flue instruments was for the first time proposed by Dünnwald (1979). He built a special organ-pipe, whose top-plate was attached to a spring. This system behaved like a harmonic oscillator coupled to the air column in the pipe. The splitting of a resonance of the pipe was observed in his experiment.

4.2.4 Resonant frequency, mass and damping coefficient of the membrane

In principle, the membrane's could be determined by its mass and effective spring constant, which could be calculated from the membrane's static area elastic modulus. In practice, this is difficult because of the involvement of wrinkles in the membrane and its fragility. Further, the damping coefficient requires an AC measurement. I therefore measure membrane's damping coefficient and resonant frequency acoustically.

The behavior of a sinusoidally driven membrane is characterized by its displacement response, defined by

$$Displacement\ response \equiv \frac{x}{x_s} = \frac{m\omega_m^2}{\omega R + jm(\omega^2 - \omega_m^2)} = \frac{m\omega_m^2}{\omega Z_{mechanical}} \tag{4.17}$$

where $\xi_s = F/m\omega_m^2$ is defined as the static displacement of the piston produced by a constant force F.

In the measurements of input impedance curves under various fingerings of the *dizi*, the membrane is mounted directly on its membrane hole. In order to obtain the displacement response curve of the membrane mounted on the *dizi*, the impedance on the membrane hole is measured with all other holes of the *dizi* closed. The acoustic impedance of the membrane coupled to a stopped pipe is calculated as follows.

Suppose that the two ends of the pipe in Fig. 4.4 are closed. The flow out of the topside of the window is

$$U_{top} = -U_1 - U_2 \tag{4.18}$$

Referring to the impedance a stopped pipe, we obtain

$$U_1 + U_2 = -\frac{pj}{Z_0 \cot kL_1} - \frac{pj}{Z_0 \cot kL_2} \tag{4.19}$$

$$\Rightarrow U_{top} = \frac{jp}{Z_0}\left(\tan kL_1 + \tan kL_2\right) \tag{4.20}$$

The total force on the membrane equals the acoustic force pS_m plus the mechanical force $Z_{acoustic}U_{top}S_m$. This yields the impedance at the membrane hole.

$$Z_{top} = \frac{-jZ_0}{\tan kL_1 + \tan kL_2} + Z_{acoustic} \tag{4.21}$$

The acoustic impedance of the membrane is altered by an imaginary term, which represents its coupling to the pipe. With the measurements of Z_{top}, L_1 and L_2, the values of m, ω_m and R can be estimated by fitting the theoretical curve of Z_{top} given by Eq. (4.21) to the experimental data.

4.3 Experimental study

4.3.1 Experimental setup

The technique for impedance measurements used for the present study is in the Music Acoustics Laboratory in the School of Physics at the University of New South Wales. It is a development of a version originally designed in to operate very rapidly for measurements of the human vocal tract (Epps *et al.* 1997). It was used for studies on the Boehm flute, the classical flute, and the baroque flute (Wolfe *et al.* 2001a, Wolfe *et al.* 2001b). A detailed description of the experimental setup can be found in these papers. It is described briefly here.

A waveform with the desired spectrum is synthesized by a computer and output via a 16 bit AD card and an amplifier to a concentric pair of loudspeakers. These are matched via an exponential horn to an attenuator, whose cross section is a narrow annulus, which attenuates standing waves in the attenuator below measurable levels, and which gives it a high output impedance (Z_a = 155 MPa.s.m^{-3} or 155 MΩ) and thus provides a nearly ideal source of acoustic current.

For calibration, the microphone at the output measures the impedance of the semi-infinite waveguide (R_{ref} = 8.5 MΩ), a stainless pipe with length of 42 m, in parallel with Z_a. This spectrum $p(f)$ includes the frequency dependence of amplifiers, speakers, horn, attenuator and microphone. A new electrical spectrum $V_{cal}(f)$, with Fourier components proportional to $1/p(f)$, is synthesized and output to the calibration load. This now produces a signal $p_{cal}(f)$ which, to the precision of the 16 bit card, is independent of frequency. $V_{cal}(f)$ is output, and the spectrum $p_{meas}(f)$ measured is now that of the measured load in parallel with the attenuator. The conductance of the latter is subtracted from the measured admittance to give $Z(f)$.

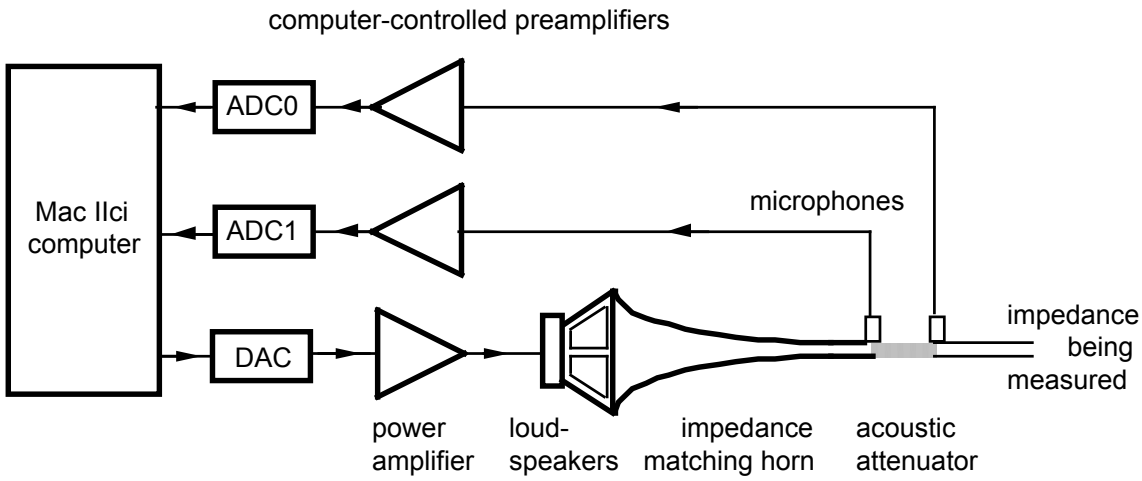


Figure 4.7: *Diagram of gear for input impedance measurements.*

4.3.2 Materials

In order to obtain meaningful impedance measurements above 2 kHz, a stainless pipe with a membrane hole in the wall at the midpoint was used. This simple, open pipe was designed to avoid the cut-off frequency of the *dizi* due to the finger holes and four end holes, thus extending the frequency range of impedance measurements to 6 kHz. Its length is 598 mm and its inner radius is 3.9 mm. A membrane hole with radius 2 mm was bored at its midpoint. This hole was covered by a curved plate with a membrane hole, as illustrated in Fig. 4.8(a).

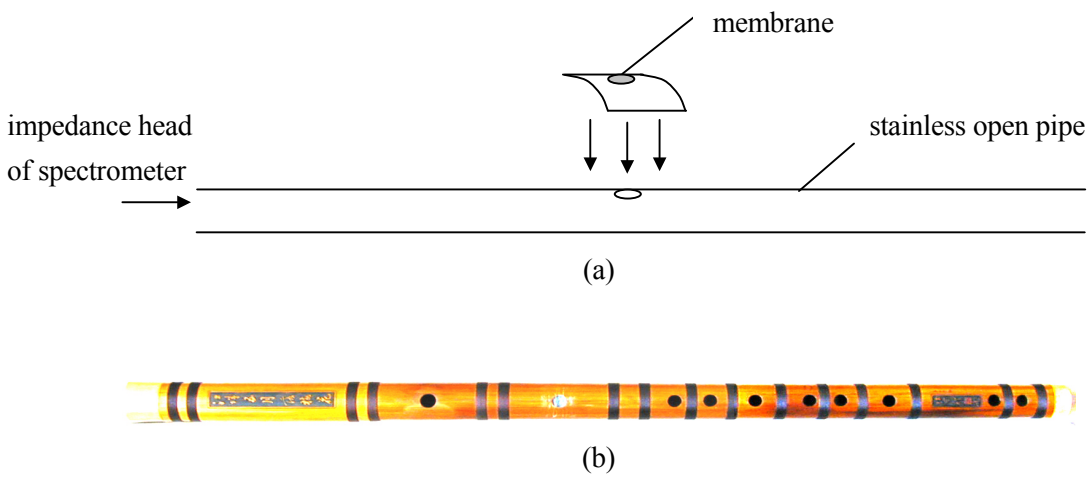


Figure 4.8: *Two pipe-membrane systems used in impedance measurements. (a) Simple open pipe and a curved plate with a membrane hole. (b) Qudi in C.*

Within various sizes of the *dizi*, the *qudi* in C was chosen for impedance measurements, whose photo is shown in Fig. 4.8(b). Its inner radius is 8 mm and the wall thickness is 5 mm. Other physical dimensions are presented in Table 4.1.

	hole length	hole width	Bore diameter	Position
Top end			16.0	-215.0
cork				-11.5
embouchure hole	10.3	9.4		0
membrane	10.0	8.5		89.5
finger hole 1	9.3	8.2		168
finger hole 2	9.3	8.2		193.0
finger hole 3	9.2	8.2		225.0
finger hole 4	9.2	8.2		262.5
finger hole 5	9.2	8.4		285.0
finger hole 5	9.6	8.5		320.0
end hole 1	9.4	8.2		380.0
end hole 2	9.2	8.1		381.0
end hole 3	9.4	8.4		397.5
end hole 4	9.4	8.4		419.0
bottom end			17.4	456.0

Table 4.1: *Physical dimensions (in mm) of the qudi in C used for this study. The mouth hole, the finger holes, the membrane hole and the end holes are all approximately elliptical in shape, with the major axis (the length) parallel to the axis of the instrument.*

4.3.3 Experimental procedure

4.3.3.1 Part one: Simple open pipe with a membrane in its wall at the midpoint

The input impedance of the simple open pipe with a membrane in the wall at the midpoint was measured. The upper frequency limit is 6 kHz and the frequency spacing was 5 Hz. The input impedance curves were measured three times. Every time various values of f_m were obtained by manipulating the membrane's tension and the number of wrinkles in it.

It was observed in experiments that the membrane's tension was a strong function of humidity and of the geometry of its mounting, and it varied with time. Measuring its resonant frequency therefore posed procedural problems. In one series of measurements, a membrane was mounted on a plate that could be affixed to the *dizi*, but could also be affixed directly to the impedance head of the spectrometer. Using this system, the impedance of the membrane on the plate was measured alone, and then I affixed the membrane plate, measuring the input impedance of the instrument. Finally, the plate was removed and the impedance of the membrane was measured again. Thereby the value of f_m and the membrane's stability could be estimated.

4.3.3.2 Part two: *Dizi*

The input impedance of the *qudi* in C was measured with twelve standard fingerings. The upper

frequency limit was 4 kHz and the frequency spacing was 2.5 Hz.

In the first series of measurements the membrane was replaced by a rigid wall. In the second series of measurements a membrane was sealed on the membrane hole. As in the part one experiment, the membrane's impedance was measured before and after the input impedance measurements. It was performed by measuring the impedance on the membrane hole with all holes of the *dizi* closed, including the four end holes, the mouth hole and the end. This relatively simple geometry made possible to estimate the acoustic impedance of the membrane by Eq. (4.21).

Finally, the input impedance of the *qudi* was measured with all finger holes and end holes closed. Under this special fingering the instrument was a simple open pipe.

4.4 Results and discussion

4.4.1 Resonance shifts and admittance reductions

The input impedance of a simple open pipe was measured for demonstrating the performance of the impedance measurement technique. Fig. 4.9 shows its impedance curve in logarithmic scale. It can be seen that the experimental data match the theoretical curve fairly well. Here the radiation impedance $Z_{radiation}$ is taken into account by referring to the acoustic resistance for a flanged piston calculated by Olson (1957).

$$Z_{radiation} = Z_0\left[\left(\frac{k^2a^2}{2} - \frac{k^4a^4}{12} + \ldots\right) + \frac{j}{\rho}\left(\frac{8ka}{3} - \frac{32k^3a^3}{45} + \ldots\right)\right] \tag{4.22}$$

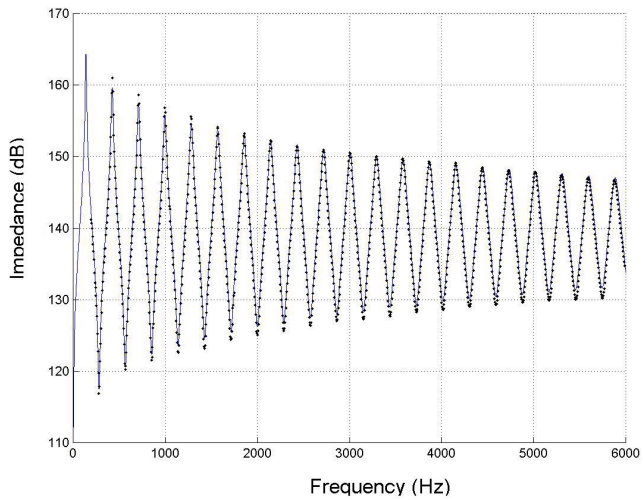


Figure 4.9: *Input impedance for an open pipe. Inner radius r_0 = 3.92 mm. Length 2L = 598 mm. Experimental data (·) and the theoretical curve (—) are compared.*

Fig. 4.10 compares the displacement response curves of the membrane measured before and after the input impedance measurement of the simple open pipe with a membrane. As can be seen, agreement with an ideal forced damped oscillator is good at low frequencies, whereas the performance of the piston model above 3 kHz is poorer.

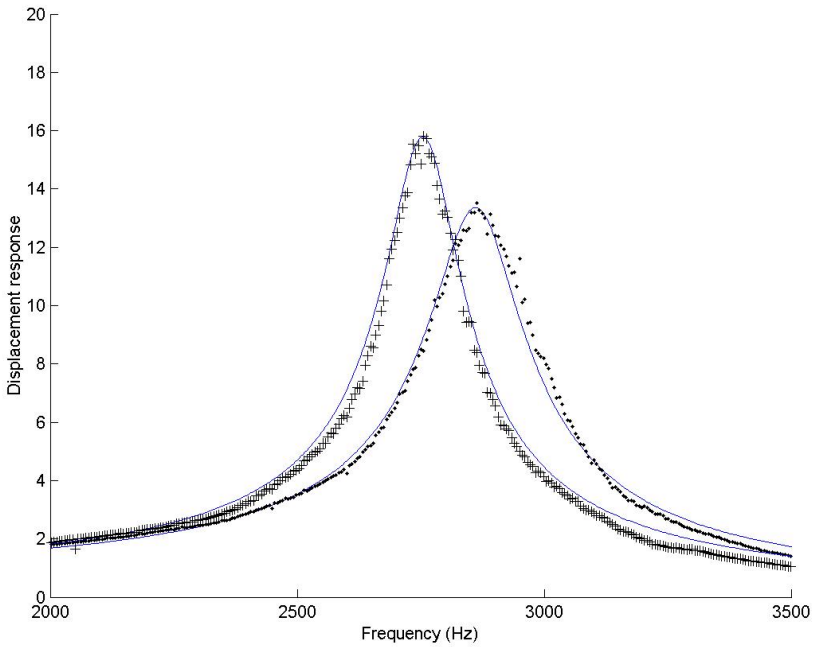


Figure 4.10: *Displacement response curves for the membrane. Experimental data (' · ' and '+') and the theoretical curves (—) are compared. Data measured before the input impedance measurement (·). Mass of the membrane m = 6´ 10^{-7} kg. Damping coefficient R=7´ 10^{-4}sec^{-1}kg. Resonant frequency f_m = 3000 Hz. Data measured before the input impedance measurement (+). Mass of the membrane m = 6´ 10^{-7} kg. Damping coefficient R = 5.7´ 10^{-4}sec^{-1} kg. Resonant frequency f_m = 2750 Hz.*

As can be noticed in Fig. 4.10, the resonant frequency and the damping coefficient of the membrane changed after the input impedance measurement of the pipe-membrane system. This reflects the instability of the membrane. As the membrane's tension appears to vary with time, the measured resonant frequency of the membrane can vary by up to 15% between the first and second measurements. Because of the membrane's unstable properties, I decided to determine the resonant frequency directly from the input impedance measurements of the pipe-membrane system by curve-fitting.

Experimental results of the input impedance measurements on the pipe-membrane system are shown in Figs. 4.11–4.13. The bore resonances are represented as minima in the input impedance curves, which are plotted in logarithm scale. These three curves are measured with different f_m. For Fig. 4.11, the value of f_m is estimated as 2.7 kHz by fitting the impedance curve to the theoretical curve. The values of f_m are estimated as 3.9 kHz and 4.6 kHz in Figs. 4.12 and 4.13, respectively.

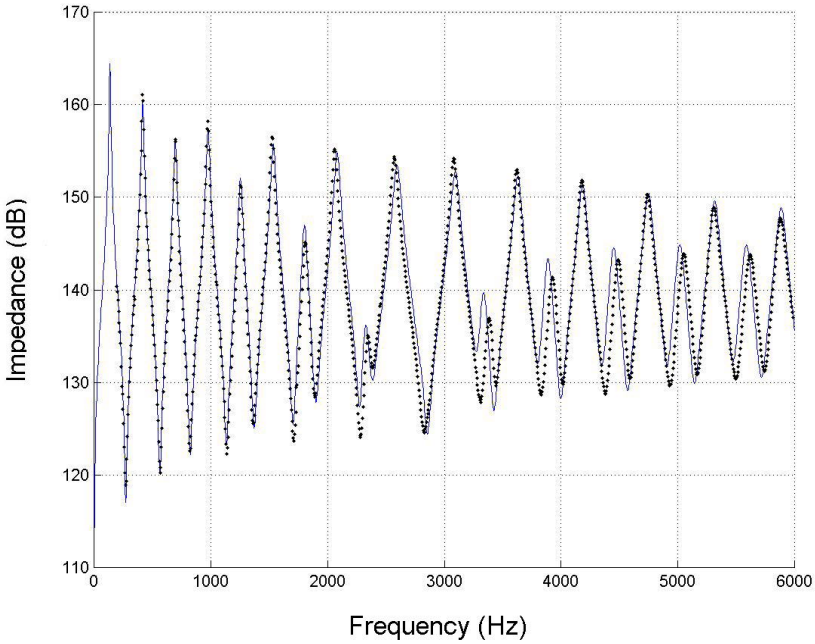


Figure 4.11: *Input impedance for an open pipe with a membrane on the hole in wall at its midpoint. Mass of the membrane m = 5×10^{-7} kg. Damping coefficient R = 6×10^{-4} sec^{-1}kg. Resonant frequency f_m = 2700 Hz. Comparison between the experimental data (·) and the theoretical curve (—).*

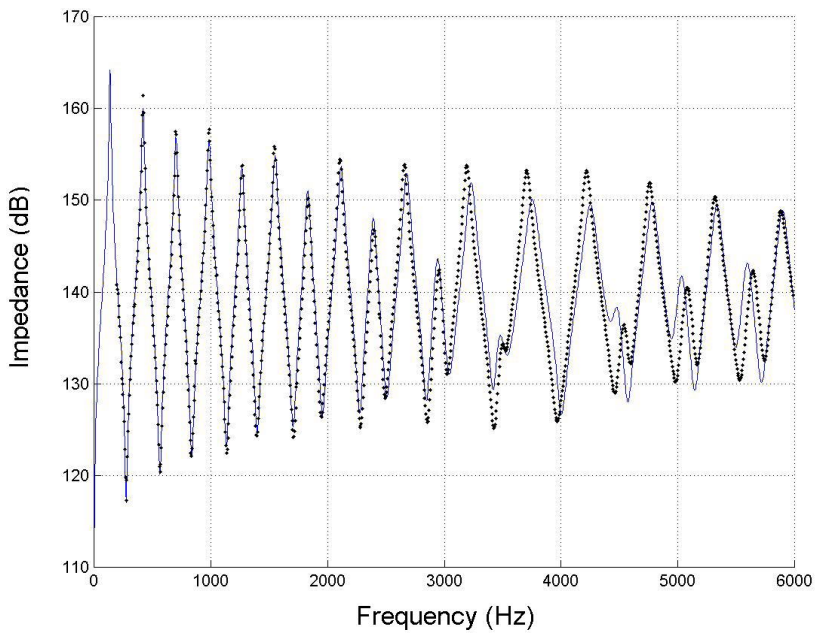


Figure 4.12: *Input impedance for an open open pipe with a membrane on the hole in wall at its midpoint. Mass of the membrane m = 4.7×10^{-7} kg. Damping coefficient R = 1.2×10^{-3} sec^{-1}kg. Resonant frequency f_m = 3900 Hz. Comparison between the experimental data (·) and the theoretical curve (—).*

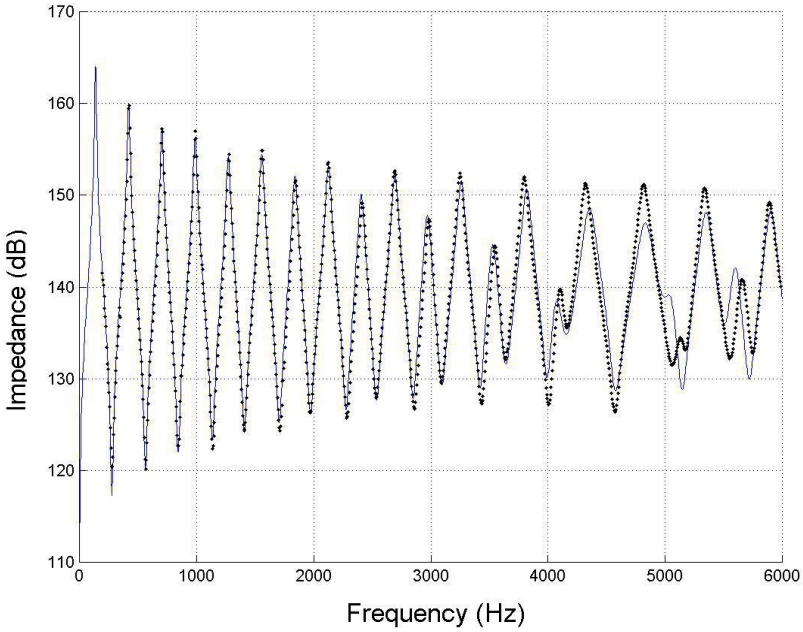


Figure 4.13: *Input impedance for an open pipe with a membrane on the hole in wall at its midpoint. Mass of the membrane m = 6×10^{-7} kg. Damping coefficient R = 2×10^{-3} sec^{-1}kg. Resonant frequency f_m = 4600 Hz. Comparison between the experimental data (·) and the theoretical curve (—).*

Resonance shifts and admittance reductions at odd-numbered resonances predicted by the piston model are demonstrated by these experimental results. The theoretical curves and the experimental data match pretty well at frequencies below f_m. Above f_m, resonance shifts are underestimated and admittance reductions at odd-numbered resonances are overestimated. The discrepancies between our model and experimental results demand further studies. However, the impedance curve at and above f_m is not important for *dizi* sound production, as its tone range is always below f_m.

The agreement between the theoretical curves and experimental data is less perfect at the even numbered resonances. Below f_m, admittance of the even-numbered resonances is underestimated, whereas it is overestimated above f_m. Comparing Fig. 4.11 with Fig. 4.9, we observe that the membrane with f_m = 2.7 kHz reduces impedance of the 6th bore resonance. This unexpected effect of the membrane on the even-numbered resonances – possibly of practice importance as this resonance lies in the *dizi* tone range – might stem from the excitation of the (11) mode of the membrane due to pressure gradient at a pressure node [Fig. 4.14(a)]. Experimental data suggest that below f_m the (11) mode vibration seems to induce an in-phase flow wave, whereas it seems to induce an in-opposite-phase flow wave above f_m [Fig. 4.14(b)]. However, it remains to be verified whether the (11) mode of the membrane plays a role in the frequency region around f_m.

Figure 4.14: *Effect of the (11) mode of the membrane on the even-numbered resonances of the pipe. (a) (11) mode of the membrane. (b) Air flow in the pipe induced by this mode.*

4.4.2 Impedance correlates of performance practice

The impedance measurements on the membrane hole of the *qudi* in C are presented in Fig. 4.15, which gives the displacement response curve of the membrane coupled to the stopped *dizi* pipe. The resonant frequency of the membrane is estimated as 2.3 kHz by curve fitting. When it was measured after the impedance measurements with twelve fingerings of the *dizi*, however, it seemed to increase to 3 Hz. Regarding the membrane's instability, it should be noted that the impedance curves of all fingerings were not measured for the same state of the membrane.

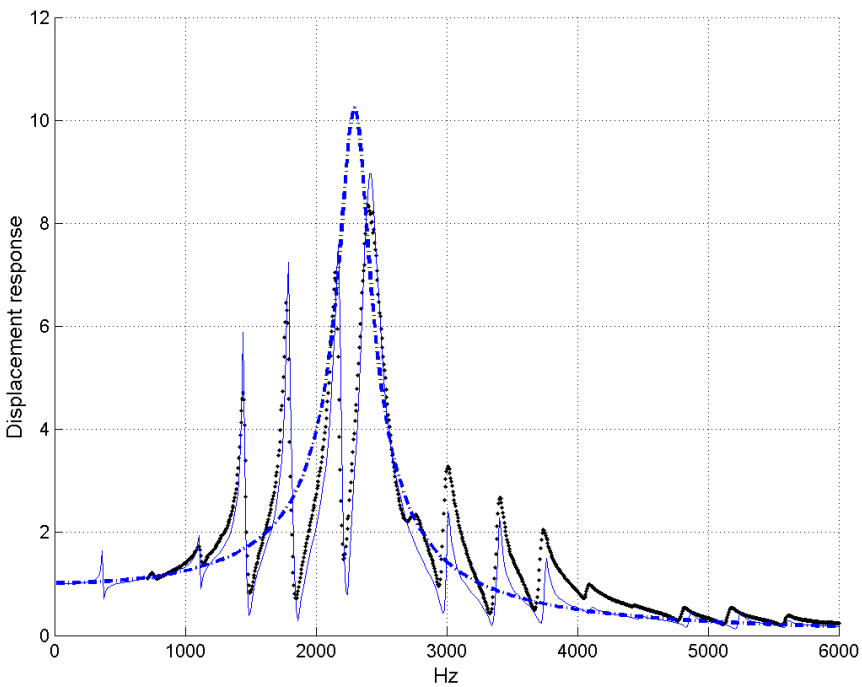


Figure 4.15: *Displacement response curve at the membrane hole of the dizi. All holes except the membrane hole were closed. The experimental data (·) and the theoretical curve (—) calculated by Eq. (4.22) are compared. The latter gives the theoretical response curve (— ·) of the membrane without coupling to a pipe. Membrane mass* $m = 1.6\times10^{-6}$ *kg. Damping coefficient* $R = 2.3\times10^{-3}$ sec^{-1}*kg. Resonant frequency* $f_m = 2.3$ *kHz.*

Based on the input impedance measurements of the *qudi* in C, Fig. 4.16 displays the resonance shifts for *dizi* notes. As can be seen, most of the bore resonances of the *dizi* are lowered by the membrane. This successfully explains pitch decreases of *dizi* tones caused by the membrane. As expected, resonance shifts are especially significant for E5, F5, G5, B6 and C7. For these tones the membrane is located near a pressure antinode of the fundamental wave in the pipe.

For the first and second octaves, resonance shifts does not exceed 20 Hz. But for *dizi* notes in the third octave (fifth register), whose fundamental frequencies are close to f_m, resonance shifts become significant. But the corresponding pitch shifts do not exceed 60 cents.

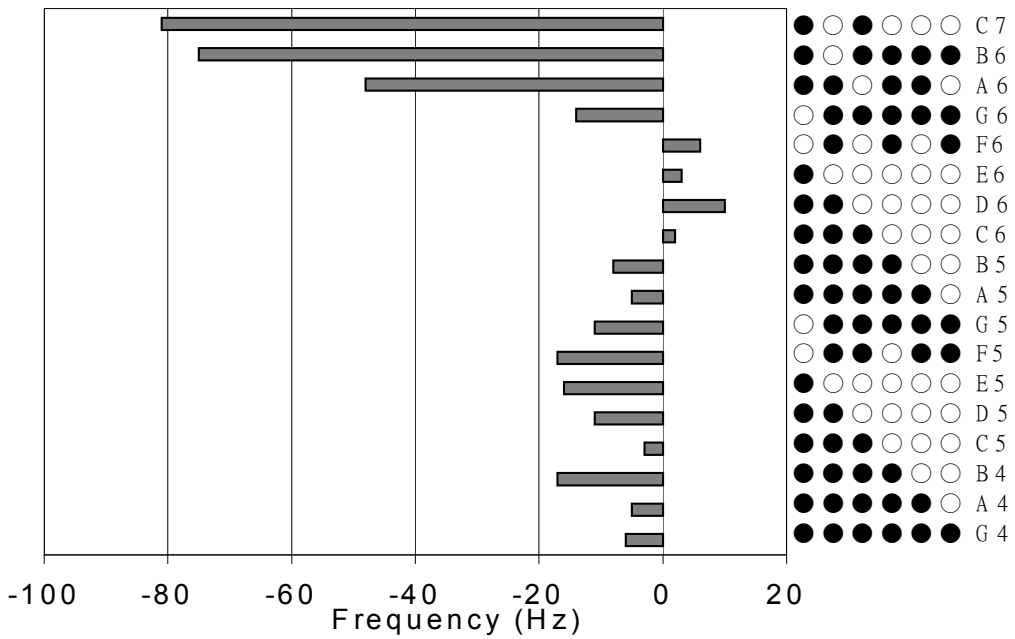


Figure 4.16: *Resonance shifts for dizi notes caused by the membrane. Except for the four notes in the fourth register (C6, D6, E6, F6), pitches are lowered by the membrane.*

The upward resonance shifts at C6, D6, E6 and F6 are not predicted by the piston model. As the membrane is approximately locate at the pressure node of the fundamental wave, these resonance shifts might be relevant to the effect of the membrane on the even-numbered bore resonances mentioned in section 4.4.1. However, the corresponding pitch shifts are small (< 10 cents).

According to the impedance measurements, resonance shifts caused by the membrane can distort the scale of the *dizi*. Here I do not intend to compare it with the tuning systems used in Western music. Nevertheless, it is important to examine the most basic interval: the octave. Fig. 4.17 compares the octaves of *dizi* tones with and without membrane vibration. When the membrane hole was covered by a rigid material, input impedance measurements showed that all octaves were slightly larger than the perfect octave. This octave stretching is exaggerated by the membrane for all notes in the first and second octaves.

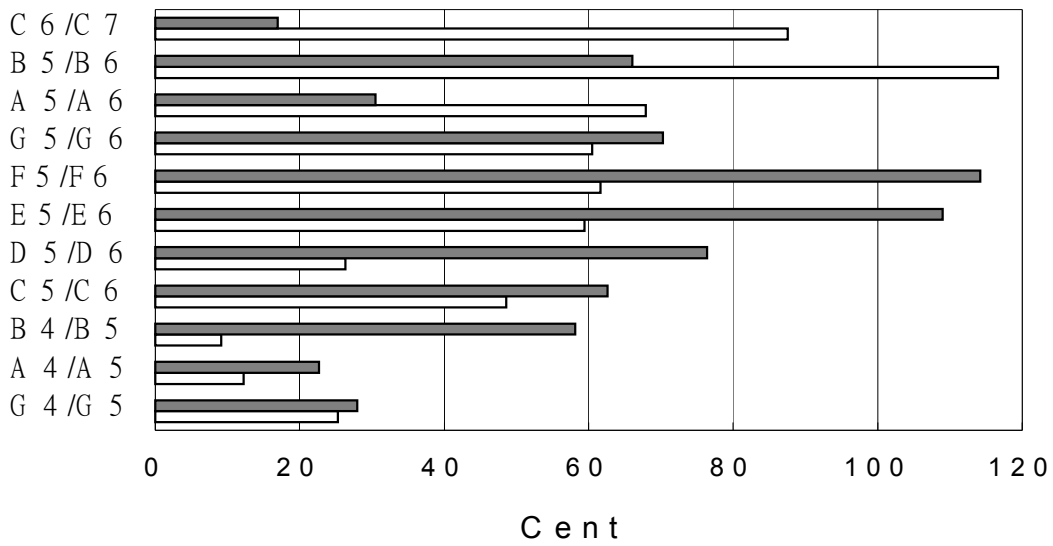


Figure 4.17: *Strechted octaves in the dizi exceeding to the perfect octave calculated by input impedance measurements. Comparison between the membrane-less flute (□) and the membrane flute (■).*

As can be noted in Fig. 4.17, the octave stretching is most significant for E5/E6 and F5/F6. E5 and E6 are played with the same fingering ●○○○○○, but their interval is stretched to a perfect octave plus a semitone when the membrane is present. This effect can be explained in terms of the membrane's location. As the membrane is located at a pressure antinode of the fundamental wave of E5, the corresponding bore resonance is shifted by the membrane. On the other hand, the membrane is located at a pressure node of the fundamental wave of E6 and the corresponding bore resonance is not affected by the membrane. Hence, the membrane results in an interval stretching for E5/E6. According to the impedance measurements, the membrane enlarged the interval of E5/E6 from 1259 cents to 1309 cents.

Although the impedance measurements show the octave-stretching effect, it should be noted that pitches of flue instruments are not solely determined by the minima of the impedance curve. Skillful players can adjust the intonation of the *dizi* by manipulating the embouchure.

In practice, these stretched octaves do not lead to serious intonation problems, but can arise timbre problems in *dizi* tones with subharmonics (octave-multiphonics). As mentioned in section 3.4.3, upper notes of the third register can have widened spectral lines of subharmonics. Such *dizi* tones are produced with the excitement of the first and second modes of the pipe. Because the octave is stretched, the coexistence of the two notes (for instance, B4/B5 or C5/C6 in the *qudi* in C) will lead to spectral-line-widening for subharmonics. This effect will be further discussed in section 6.4

Next to resonance shifts, the membrane also affects impedance of resonances, which determine the playability of *dizi* notes. Fig. 4.18 shows the changes in impedance caused by the membrane. The membrane does not significantly affect the resonance impedance for most of the *dizi* notes. Only the

bore resonances supporting B6 and C7, whose fundamental frequencies fall close to f_m, show large impedance changes.

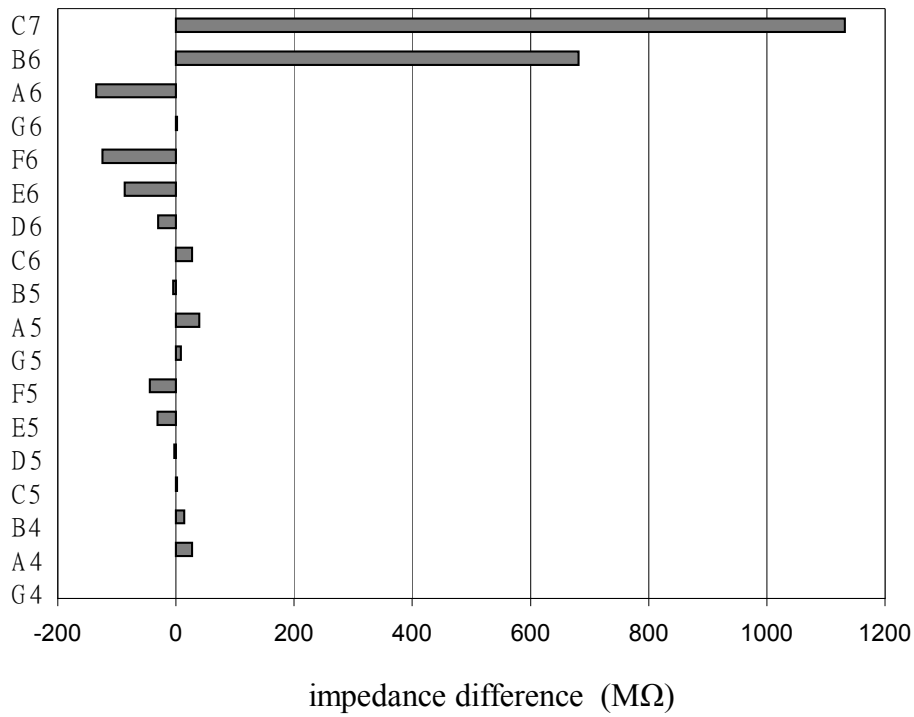


Figure 4.18: *Impedance changes of resonances corresponding to dizi notes caused by the membrane. Only the resonances of the two highest notes B6 and C7 show significant impedance increases.*

The input impedance curves for the two notes B6 and C7 are given in Figs. 4.19 and 4.20. Resonance shifts and impedance increases around 2 kHz are noticeable. As a high value of impedance implies a low playability of the *dizi* note, the impedance increases of resonance supporting B6 and C7 successfully explain that these two high notes are impossible to play unless the membrane is very thick and taut. It should be noted in Fig. 4.18 that the impedance minimum at 2.5 kHz is shifted upward by the membrane, while the one at 2 kHz is shifted downward. This implies 2 kHz < f_m < 2.5 kHz. It is consistent with the measurement of the membrane's resonant frequency f_m = 2.3 kHz shown in Fig. 4.14.

Finally, the input impedance measurements of the *dizi* with all finger-holes and end-holes closed are presented in Fig. 4.20. This special fingering avoided the cut off frequency by covering all the four end holes; the instrument becomes a simple open pipe. A resonance splitting at 2.7 kHz caused by the membrane can be observed. It recalls the wolf tone in bowed-string instruments. However, admittance at the two split resonances is too low for flute tone production and it will be even lower when the end holes are open. I therefore conclude that the wolf tone does not take place in the *dizi*.

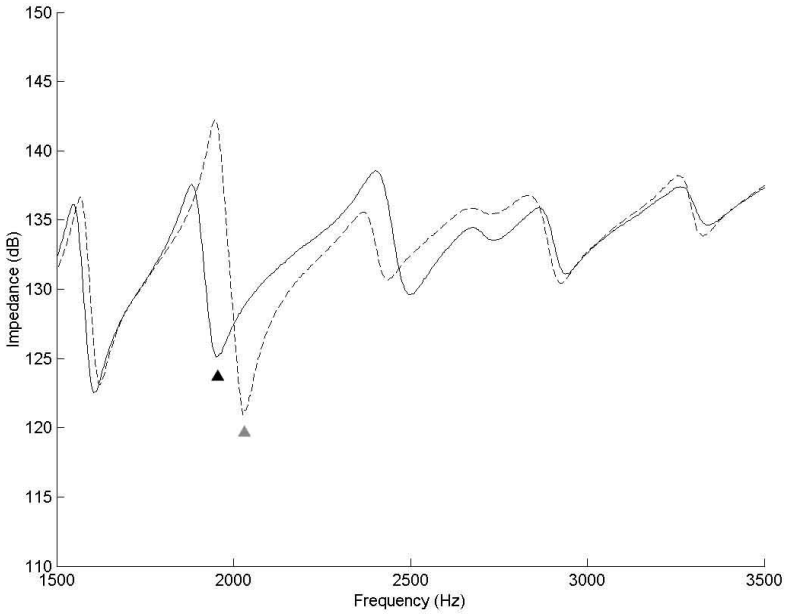


Figure 4.19: *Input impedance for B6 of the qudi in C. Fingering: ●○●●●●. Comparison between the measurements with (—) and without (- - -) the membrane vibration. The bore resonance corresponding to B6 is at 2026 Hz (marked with a grey triangle) when the membrane is replaced by a rigid wall, while it is at 1951 Hz with the vibrating membrane (marked with a black triangle).*

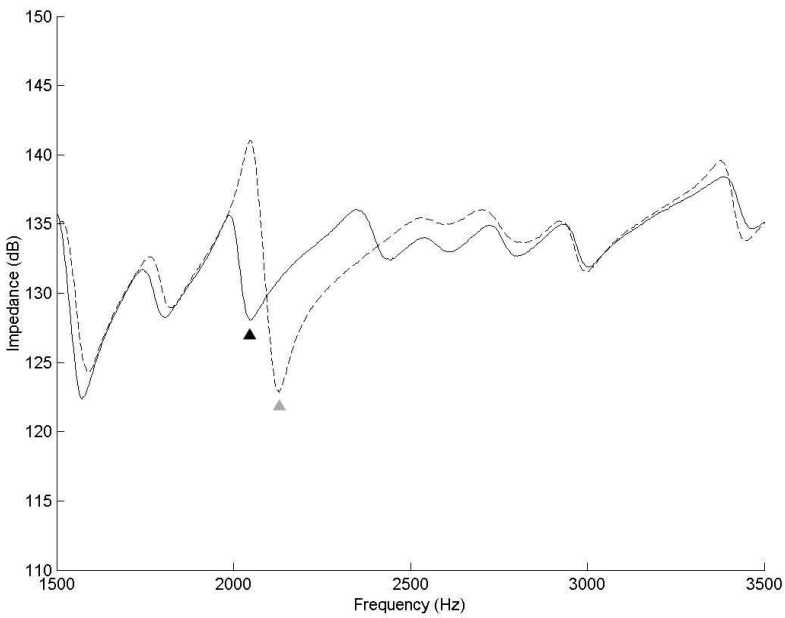


Figure 4.20: *Input impedance for C7 of the qudi in C. Fingering: ●○●○○○. Comparison between the measurements with (—) and without (- - -) the membrane vibration. The bore resonance corresponding to C7 is at 2129 Hz (marked with a grey triangle) when the membrane is replaced by a rigid wall, while it is at 2048 Hz with the vibrating membrane (marked with a black triangle).*

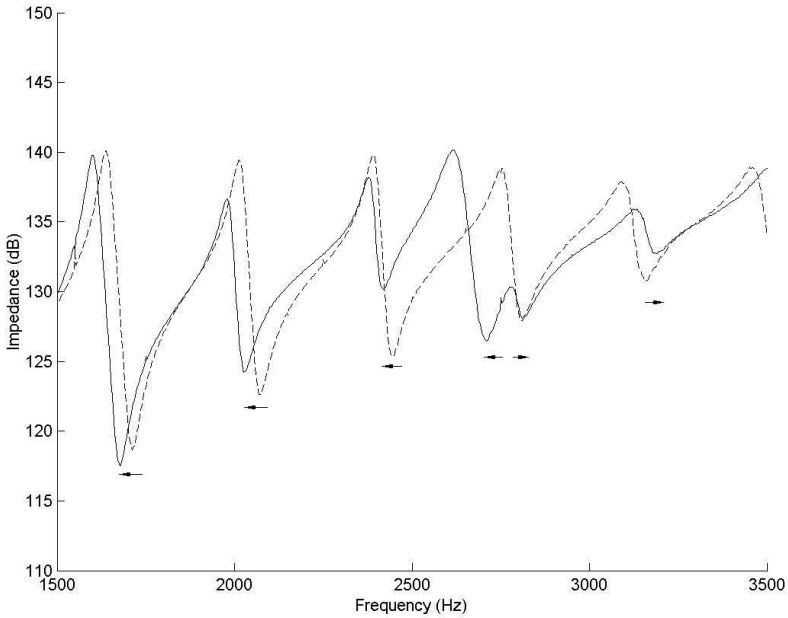


Figure 4.21. *Input impedance for the dizi with all finger-holes and end holes closed. Comparison between the measurements with (—) and without (- - -) the membrane vibration. The membrane results in a resonance splitting at 2.7 kHz.*

According to the impedance measurements in this experiment, typical values of f_m lie between 2 kHz and 3 kHz for the *qudi* and the large *dizi*. This guarantees a characteristic bright timbre due to the large-amplitude vibration of the membrane. A low value of f_m can be obtained by using a thin membrane and sealing it slackly on a larger membrane hole. On the other hand, for the *bangdi*, the Chinese piccolo, a higher value of f_m is needed for producing high notes. With a high tension in the *bangdi* membrane, f_m can be higher than 3 kHz. Some musicians suggest that the *bangdi* should be sealed by a thicker membrane (Jan 1991). This can be explained by Eqs. (4.13) and (4.14); because a larger mass of the *dizi* membrane reduces its coupling to the air column in the pipe, reductions in admittance of resonances can be moderated. However, when a *bangdi* membrane has a high value of f_m, its vibration amplitude will be small in soft or low tones as the acoustic pressure driving the membrane is small. Therefore, soft or low tones of the *bangdi* tend to be poor in higher harmonics.

The present experiments suggest that the membrane restricts the tone range of the *qudi* in C in two octaves plus two notes. Beyond this range, B6 and C7 are unplayable unless the membrane is very taut and thick. The membrane's resonant frequency f_m sets an upper limit of the dizi tone range through admittance reductions of the resonances supporting these two high notes, whose fundamental frequencies are close to f_m. Admittance reductions of bore resonances stem from the pipe's coupling to the membrane, whose damping is much larger than the wall lose in the pipe.

4.5 Conclusions

The linear behavior of the *dizi* membrane is studied by modeling it as a piston attached to a linear spring. Impedance measurements gave typical values of parameters in the piston model: the membrane's resonant frequency 2 kHz < f_m < 3 kHz, the effective acoustic mass m of the order of 10^{-6} kg, and the damping coefficient R of the order of 10^{-3} sec^{-1}kg.

In the experiment for a simple open pipe with a membrane hole at its midpoint, admittance reductions and resonance shifts at odd-numbered resonances in the frequency range below f_m were quantitatively in agreement with impedance calculations based on the linear piston model. The even-numbered resonances around f_m were also affected by the membrane. This unexpected result might be due to the excitation of the (11) vibration mode of the membrane.

The impedance measurements of the *qudi* in C show that the membrane caused significant admittance reductions for the bore resonances supporting the two notes B6 and C7, and thus the *dizi* range is restricted to about two octaves plus two notes. Analogously to a string-soundboard system, the vibration modes of the pipe-membrane resonator of the *dizi* are largely damped by the membrane when the pipe's resonances are near the membrane's resonant frequency f_m. As f_m sets an upper limit of the tone range, the membrane of the *bangdi* tends to be sealed more tautly than that of the *qudi* for producing high notes.

References

Dünnwald, H. (1979). Versuche zur Entstehung des Wolfs bei Violininstrumenten. Acustica 41, 238-245.

Epps, J., Smith, J., and Wolfe, J. (1997). A novel instrument to measure acoustic resonances of the vocal tract during speech, Measurement Science and Technology 8, 1112-1121.

Fletcher, N.H., and Rossing, T.D. (1991). The Physics of Musical Instruments. Springer-Verlag.

Gough, C.E. (1981). The theory of string resonances on musical instruments. Acustica 49, 124-141.

Güth, W. (1978). Der Wolfton beim Cello. Acustica 41, 164-176.

Jan, Y.M. (1991). *Dizi Jichu Jiaucheng*.（詹永明《笛子基礎教程》，達欣出版社）

Olson, H.F. (1957). Acoustical Engineering. Van Nostrand-Reinhold, Princeton.

Raman C.V. (1916). On the wolf-note in bowed string instruments. Phil. Mag. 32, 391-395.

Schelleng, J.C. (1963). The violin as a circuit. Journal of the Acoustical Society of America 35, 326-338.

Wolfe, J., Smith, J., Tann, J., and Fletcher, N.H. (2001a). Acoustic impedance of classical and modern flutes. Journal of Sound and Vibration 243, 127-144.

Wolfe, J., and Smith, J. (2001b). Embouchures and end effects in air-jet instruments. International Congress on Acoustics, Rome, Session 8.06, 8-9.

Chapter 5.
Non-linearity of the membrane: Wrinkling and performance practice

Abstract

The non-linear behavior and wrinkling of a *dizi* membrane, crucial for producing a "correct" *dizi* timbre, were studied theoretically and experimentally. Theoretically, a circular membrane with an isotropic initial tension behaves like a Duffing oscillator with a hardening spring (Chobotov and Binder 1964). In the case of the *dizi* membrane, its elastic property is complicated by wrinkling, which is produced by a uniaxial tensile strain. Theoretical study suggests that a wrinkled membrane has a higher initial tension ($4N_0$) than a membrane with an isotropic initial tension (N_0) for the same resonant frequency. Wrinkling reduces the cubic non-linearity of the membrane, avoiding a jump phenomenon in its response curve.

The existence of the membrane's non-linearity was demonstrated by driving it with sinusoidal acoustic waves strong enough to induce its harmonic generation. Relations between the magnitudes of the lowest three harmonics revealed a predominance of cubic non-linearity for an unwrinkled membrane, and a predominance of quadric non-linearity for a wrinkled membrane in an intensity range of the driving signals. The membrane's response curves showed the frequency-pulling effect characteristic of a hardening spring. A jump phenomenon was observed in the multi-valued response curve of a slack, unwrinkled membrane. In an impedance calculation, the membrane with a large/small jump was approximated as a linear oscillator with a large/small damping coefficient. The large damping could lead to significant admittance reductions of the bore resonances supporting notes for which the membrane is located near a pressure antinode of the fundamental wave.

5.1 Introduction

In chapter 4 the input impedance of the *dizi* was measured by exciting the instrument with acoustic signals. As the signals used were much weaker than the acoustic waves in the *dizi* pipe during

a performance, the impedance measurements only revealed the linear behavior of the passive resonator of the *dizi*. The focus of this chapter turns to the membrane's non-linearity, which is expected to explain the harmonic-rich sound of the *dizi*. If the *dizi* membrane's behavior is non-linear in a typical playing range, it will transfer the acoustic energy in the pipe from lower frequencies to higher frequencies. The purpose of the present study is to demonstrate the existence of the membrane's non-linearity using signals stronger than those in the previous impedance measurements. For the sake of simplicity, only the quadric non-linearity and cubic non-linearity are considered here.

As the tension and wrinkles in a *dizi* membrane are essential to production of the characteristic *dizi* timbre (Thrasher 2001), their influences on the membrane's non-linearity are of interest. It is known by *dizi* musicians that a very taut membrane cannot produce characteristic bright *dizi* tones. But if the membrane is too slack, the *dizi* will become difficult to control. It can be so unstable that it either does not sound, or sounds raucously under the same fingering. In other words, its persistent behavior appears to be sensitive to the initial condition – a common phenomenon in non-linear systems. In order to avoid these problems, musicians have to stretch the membrane carefully to produce a lot of lateral wrinkles and set an intermediate value of tension on them. The influences of the tension and wrinkles on the membrane's linear and non-linear properties are studied theoretically and experimentally.

The sources of the membrane's non-linearity and the effects of wrinkles in it are discussed in section 5.2. Its non-linear behaviors corresponding to the cubic and quadric non-linearity are distinguished in section 5.3. The experimental methods are addressed in section 5.4 and the results are presented in section 5.5. Section 5.6 relates the jump phenomenon associated with the cubic non-linearity to performance practice of the *dizi*.

5.2 Membrane's non-linearity and wrinkles

5.2.1 Unwrinkled membrane

5.2.1.1 Source of cubic non-linearity

In a non-linear piston model of a membrane, the involvement of cubic non-linearity corresponds to a hardening or a softening spring. A circular membrane with an isotropic tension was shown to have the property of a hardening spring. Chobotov and Binder (1964) derived the equations for rotationally symmetric motions, namely the (01) mode, of a circular membrane without damping, which was driven by acoustic signals. With the help of a perturbation procedure and the Ritz-Galerkin technique, the exact differential equations of the membrane can be written as an equivalent *Duffing equation* (see, e.g., Hagedorn 1995)

$$\ddot{q} + \omega_m^2 q + \beta q^3 = 0 , \tag{5.1}$$

with

$$\omega_m^2 = \frac{6N_0}{\rho_m h_m a^2} \tag{5.2}$$

and

$$\beta = \frac{\omega_m^2 w_0^2 (13 + 21\mu - 4\mu^2)}{30a^2 \gamma (1+\mu)}, \tag{5.3}$$

where q is the non-dimensional displacement of the membrane in (01) mode, ρ_m the membrane's density (pro volume), h_m its thickness and w_0 its central displacement. N_0 is the initial tension and μ its Poisson's ratio. They are related to the initial strain (change in area per unit of original area) γ and the Young's modulus E by the equation

$$\gamma E h_m = (1-\mu) N_0 . \tag{5.4}$$

For most material the value of μ is about 0.3. According to Eq. (5.3) the coefficient of cubic non-linearity β is positive, and hence the membrane has a hardening-spring characteristic. Furthermore, Chobotov and Binder (1964) assumed a linear damping, which was verified by the experiment for small and medium deflections of the membrane.

The linear property of this membrane can be noted by Eq. (5.2), which gives the effective spring constant

$$K_{eff} = 6N_0 \pi . \tag{5.5}$$

On the other hand, Eq. (5.3) indicates that the non-linear stiffness decreases as the initial tension increases and the displacement decreases.

$$\beta \propto \gamma^{-1} w_0^2 . \tag{5.6}$$

5.2.1.2 Source of quadratic non-linearity

The quadratic non-linearity implies an asymmetry of the membrane's potential energy between the upward and downward displacements. I propose that the global curvature of the membrane is responsible for this asymmetry.

When a membrane is sealed on a hole in the wall of the cylindrical tube, it is more or less shaped like a saddle surface [Fig. 5.1(a)]. In the case that the two curvatures at the saddle point have the same magnitude, the membrane is approximately symmetric for both sides. In the case of a *dizi* membrane, the symmetry between the upward and downward vibrations is broken, as the curvature on the plane perpendicular to the *dizi* (the *yz*-plane) is approximately zero [Fig. 5.1(b)]. This is due to a vanishing

tension in the direction paralle to the *dizi* pipe (the *y* direction).

Fig. 5.2 compares the upward and downward vibrations of a *dizi* membrane. For a pressure difference across the two sides of the membrane, it will deflect more when exerted by an upward force [Fig. 5.2(a)], and deflect less when exerted by a downward force [Fig. 5.2(b)]. This symmetry breaking is due to a larger area increase of the membrane for downward deflections. The quadratic non-linearity will vanish when the membrane is sealed on a hole in a flat wall instead of a cylindrical wall.

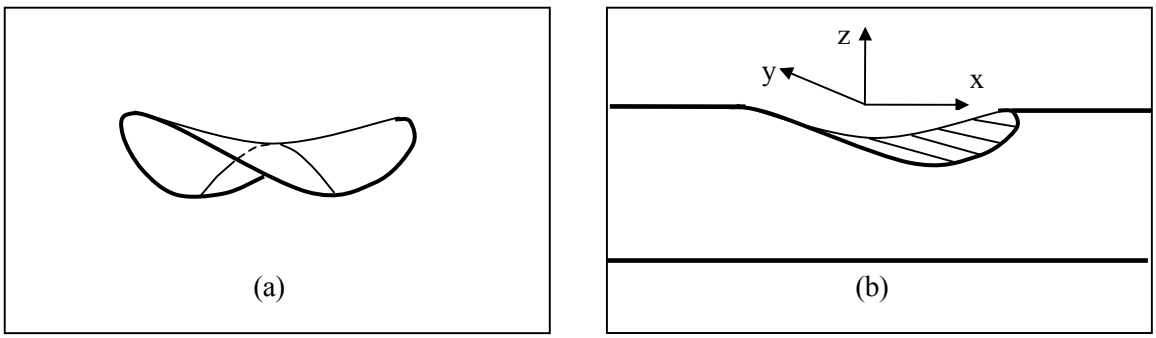


Figure 5.1: *Curved membrane sealed on a circular hole in the wall of a cylindrical tube. (a) Membrane shaped like a saddle surface. (b) Dizi membrane, which is usually sealed in a manner that its curvature in the plane perpendicular to the dizi pipe (yz-plane) is approximately zero.*

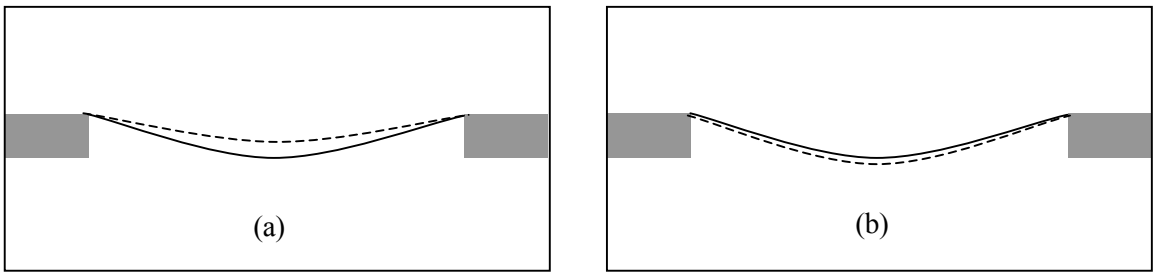
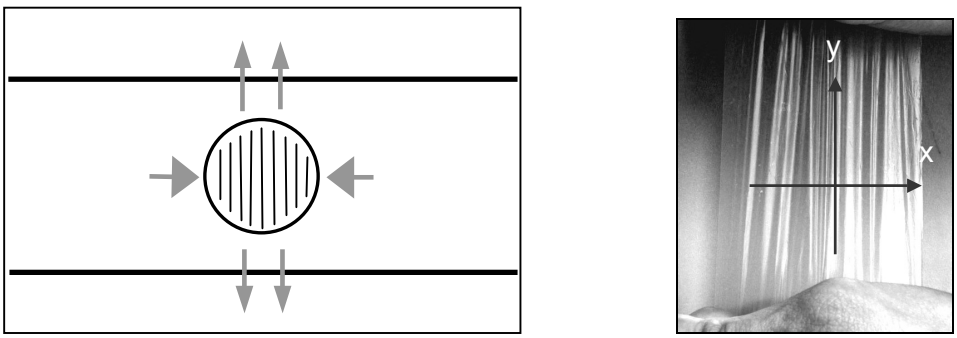


Figure 5.2: *Comparisons between the static position (solid lines) and the deformed position (dashed lines) of a dizi membrane. (a) Upward vibration. (b) Downward vibration.*

5.2.2 Wrinkled membrane

5.2.2.1 Energy balance in static wrinkling

As mentioned in section 3.2, wrinkles in the *dizi* membrane play a crucial role in sound production. Wrinkles in the membrane are generated by imposing a stretching perpendicular to the *dizi* pipe and a compressive displacement parallel to the *dizi* pipe to the membrane [Fig. 5.3(a)]. They allow the membrane to vibrate in such a manner that the characteristic bright, sweet the *dizi* tones can be produced.



(a) (b)

Figure 5.3: *Wrinkles due to external forces and/or geometrical constraints. (a) Wrinkles in the dizi membrane, which are generated by imposing a stretching perpendicular to the dizi pipe and a compressive displacement parallel to the dizi pipe. (b) Wrinkles in a polyethylene sheet under a uniaxial tensile strain.*

The behavior of a wrinkled membrane is governed by a set of non-linear partial differential equations, known as Föppl-von Karman equations (Landau and Lifshitz 1986), which cannot be solved analytically. Cerla and Mahadevan (2003) recently made progress in the geometry and physics of static wrinkling using scaling and asymptotic arguments. They successful quantified the wavelength of wrinkling in a stretched elastic sheet due to external forces and/or geometrical constraints. It is worthy to mention their results here.

Consider a stretched, slender elastic sheet of thickness h_m, width $2W$, length $2L$ made of a material with Youngs modulus E and Poisson's ratio μ under a stretching strain γ in the y direction [Fig. 5.3(b)]. Assume that the out-of-plane displacement of the initially flat sheet is $\zeta(x,y)$, where $-W < x < W$ and $-L < y < L$. The bending energy U_b and the stretching energy U_s are respectively dominated in the y and x directions:

$$U_s = \frac{1}{2}\int N_0(x)\left(\partial_y \zeta\right)^2 dA \tag{5.7}$$

and

$$U_b = \frac{\kappa}{2}\int\left(\partial_x^{\,2}\zeta\right)^2 dA\,, \tag{5.8}$$

where N_0 is the tension and κ is the bending stiffness, given by

$$\kappa = \frac{Eh_m^{\,3}}{12(1-\mu^2)}\,. \tag{5.9}$$

As the sheet wrinkles in the x direction under the action of a small compressive stress, it satisfies

the condition of inextensibility

$$\int_{-W}^{W}\left[\frac{1}{2}(\partial_x z)^2-\frac{\Delta(y)}{2W}\right]dx=0\,, \tag{5.10}$$

where $\Delta(y)$ is the imposed compressive displacement.

The selection of the wavelength of wrinkling was shown to be determined by a compromise minimizing the bending energy and the stretching energy subject to the geometry constraint of inextensibility. The form of U_b in Eq. (5.7) makes it transparent that the total energy increases rapidly for short wavelength. On the other hand, the stretching energy also increases with long wavelengths due to the increase in the amplitude, which stems from the constraint of inextensibility Eq. (5.9). In either cases, the balance between the bending and stretching energies leads to the selection of wrinkles of an intermediate wavelength.

5.2.2.2 Wrinkling correlates of cubic non-linearity

Cerla and Mahadevan's (2003) theoretical study on wrinkled sheets clamped on two sides provides insight into the dynamical behavior of a wrinkled *dizi* membrane, which is sealed on a circular hole. In the following discussion, the curvature of the *dizi* tube is ignored; the boundary of the *dizi* membrane is assumed to lie in the *xy*-plane.

The predominance of the stretching energy in the *y* direction allows me to separate the two parameters *x* and *y* of the out-of-plane displacement of the *dizi* membrane; the membrane is divided into a series of strings parallel to the *y*-axis, as illustrated in Fig. 5.3(a). The motion of the wrinkled membrane can be described in terms of these strings, which are coupled together mainly by the bending energy in the *x* direction. If the thickness of the membrane h_m is so small that the bending energy can be ignored, wrinkling can lead to a decoupling between these strings. Note that the geometrical constraint due to excess membrane along the *x*-axis allows the amplitude of wrinkling to adjust during vibration [Fig. 5.4(b)]. When the vibration amplitude is large enough, however, the wrinkling amplitude vanishes and the geometrical constraint Eq. (5.10) is no longer satisfied. In such situations, these strings are coupled together by a non-ignorable stretching energy in the *x* direction and the membrane shows a hardening spring characteristc.

The effective spring constant of a wrinkled membrane in medium-amplitude vibrations can be calculated. Under a small pressure difference *p* across the two sides of a wrinkled membrane, these decoupled strings are parabolic curves because

$$N_0(x)\frac{\partial^2 z\,(x,y)}{\partial y^2}=p\,. \tag{5.11}$$

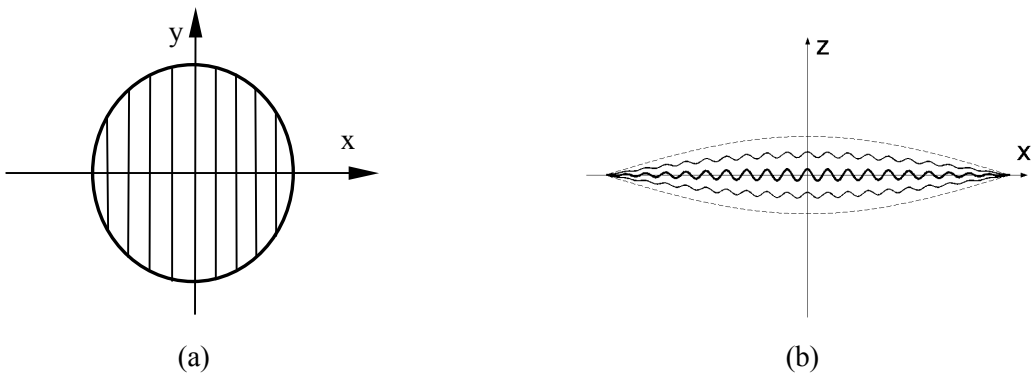


Figure 5.4: *Decoupling of vibrating strings in a wrinkled dizi membrane. (a) Dizi membrane divided into strings parallel to the y-axis. (b) Cross-section of the membrane in the xz-plane. The static position (thick line) is compared to small deformations (thin line) and large deformations (dashed line). Note that the wrinkling amplitude varies while the membrane is vibrating, so that the constraint of inextensibility Eq. (5.9) is satisfied and the strings are decoupled. For larger deformations, the wrinkling amplitude vanishes and these strings are coupled together by a stretching energy in the x direction.*

Here the geometry of wrinkles is ignored by assuming that the amplitude and wavelength of wrinkles are much smaller than the size of the membrane. For a uniform tension N_0, it is reasonable to approximate the shape of an acoustically driven *dizi* membrane as

$$z(x,y,t) = h(t)(a^2 - y^2 - x^2)\,. \tag{5.12}$$

The effective spring constant K_{eff} of a wrinkled *dizi* membrane stretched in the y direction with tension N_0 can be obtained by equating its stretching energy to the effective potential energy U_p of a piston, which is proportional to the square of the effective displacement of the membrane ξ:

$$x = \int \frac{z(x,y)}{pa^2} dA \tag{5.13}$$

$$U_p = \frac{K_{eff}}{2} x^2\,. \tag{5.14}$$

This potential energy must be equal to the stretching energy given by Eq. (5.7). This results in

$$\frac{K_{eff}}{2}\left(\int_{-a}^{a}\int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} \frac{z(x,y)}{pa^2} dydx\right)^2 = \frac{N_0}{2}\int_{-a}^{a}\int_{-\sqrt{a^2-y^2}}^{\sqrt{a^2-y^2}} (\partial_x z)^2 dydx\,. \tag{5.15}$$

Substituting Eq. (5.12) to Eq. (5.15) yields

$$\frac{K_{eff}}{2(\pi a^2)^2}\left(\int_{-a}^{a}\frac{4\eta(a^2-x^2)^{3/2}}{3}dx\right)^2=\frac{N_0}{2}\int_{-a}^{a}\frac{2\eta^2(a^2-x^2)^{3/2}}{3}dx \tag{5.16}$$

$$\Rightarrow \qquad K_{eff}=\frac{3N_0\pi}{2} \tag{5.17}$$

This is the relation between the initial tension and the effective spring constant of the non-linear piston model of a circular membrane under a uniaxial tensile strain. Referring to Eq. (5.5) for the membrane with an isotropic tension, it can be found that to get the same effective spring constant, the membrane under a uniaxial tensile strain must has a tension considerably larger than that of the unwrinkled membrane with an isotropic tension.

For the same spring constant or resonant frequency, a wrinkled *dizi* membrane under a uniaxial tensile strain has a cubic non-linearity lower than an unwrinkled membrane with an isotropic tension. First, a low initial strain in the unwrinkled membrane will lead to a large cubic non-linearity because its tension will significantly vary in large-amplitude vibrations; it is not fully stretched. This makes the effective spring of the membrane fairly hardened. Second, wrinkles largely reduce the coupling of the strings. Their behavior is fairly linear until the constraint of inextensibility Eq. (5.10) is no longer satisfied for large deformations and these strings are coupled by a stretching energy in the x direction.

In summary, the main effect of wrinkles in the *dizi* membrane is to make it behave more linearly by reducing the cubic non-linearity. If the *dizi* membrane is unwrinkled due to an isotropic initial tension, the spring constant will be too high for a high initial strain so that the membrane hardly vibrates, or, the cubic non-linearity will be too high for a low initial strain. In the latter case, a jump phenomenon in the response curve of the membrane could occur in a typical playing range of the *dizi*. This will be borne out in the experiment.

In the above discussion, the bending energy U_b given by Eq. (5.8) is ignored. For a thick membrane, U_b may play a role in the membrane's vibration by softening the spring, because U_b reaches its maximum when the effective displacement of the membrane ξ is zero. Moreover, the changing amplitude of wrinkling during vibration is likely to be the main source of damping, which may be non-linear. Finally, the quadric non-linearity due to the membrane's curvature in its static position could affect its vibration. These complications are not taken into account in the present crude model of a wrinkled membrane.

5.3 Theory

5.3.1 Cubic non-linearity and jump phenomenon

At the neighborhood of the membrane's static position, its restoring force can be expressed as a Taylor series:

$$F_{restorting} = m(\omega_0^2 x + a x^2 + b x^3 + \ldots) . \tag{5.18}$$

The first term represents the Hook's law, while the second term corresponds to the quadric non-linearity and the third term corresponds to the cubic non-linearity. In the first approximation, only the first three terms are considered. If the restoring potential is symmetric, the first non-linear term in Eq. (5.18) is the cubic term. When the membrane is driven by a sinusoidal force, its motion equation is

$$\ddot{x} + r\dot{x} + \omega_0^2 x + b x^3 = \frac{G_0}{m}\cos(\omega t + \phi_0) . \tag{5.19}$$

If β is small, it is easy to demonstrate that the solution of the Duffing equation is of the following form:

$$x = B_1 \cos\omega t + B_3 \cos(3\omega t + \phi_3) + B_5 \cos(5\omega t + \phi_5) + B_7 \cos(7\omega t + \phi_7) + \ldots , \tag{5.20}$$

where the coefficients are all real numbers. Substituting Eq. (5.20) to Eq. (5.19) and collecting terms of $\cos\omega t$ and $\sin\omega t$ give

$$(\omega_0^2 - \omega^2)B_1 \cos\omega t - r\omega B_1 \sin\omega t + b B_1^3 \frac{3\cos\omega t}{4} = \frac{G_0}{m}\cos(\omega t + \phi_0) . \tag{5.21}$$

The coefficients of $\cos\omega t$ and $\sin\omega t$ must vanish, so

$$(\omega_0^2 - \omega^2)B_1 + \frac{3}{4} b B_1^3 = \frac{G_0}{m}\cos\phi_0 \tag{5.22}$$

and

$$-r\omega B_1 = -\frac{G_0}{m}\sin\phi_0 . \tag{5.23}$$

Squaring both sides of the two equations and summing them give

$$\left[(\omega_0^{\,2}-\omega^2)B_1+\frac{3}{4}\beta B_1^{\,3}\right]^2+(r\omega B_1)^2=\left(\frac{G_0}{m}\right)^2. \tag{5.24}$$

The fundamental amplitude A_1 can be expressed as a function of the driving frequency by solving Eq. (5.19). Response curves of softening and hardening spring systems with various magnitudes of the driving force are showed in Fig. 5.5(a)(b). For a softening spring system, the effective spring constant and the resonance decrease as the driving force magnitude increases, while they increase for a hardening spring system. This effect is sometimes called the *frequency-pulling effect*. When the force is large enough, the response curve can be multi-valued; A_1 have three values, one is stable, one is meta-stable and one is unstable. In this frequency region, different histories of the motion lead to different persistent behaviors. If the forcing frequency varies gradually while keeping all the other parameters fixed, a *jump phenomenon* can be observed (see, e.g., Hagedorn 1995). Starting from small forcing frequency and slowly increasing, the response will suddenly exhibit a finite jump in amplitude and follow the upper path, as illustrated in Fig. 5.5(c). On the other hand, if we start with a large forcing frequency and gradually reducing, the response will again reach a critical state resulting in a sudden jump down to the small-amplitude solution, as illustrated in Fig. 5.5(d).

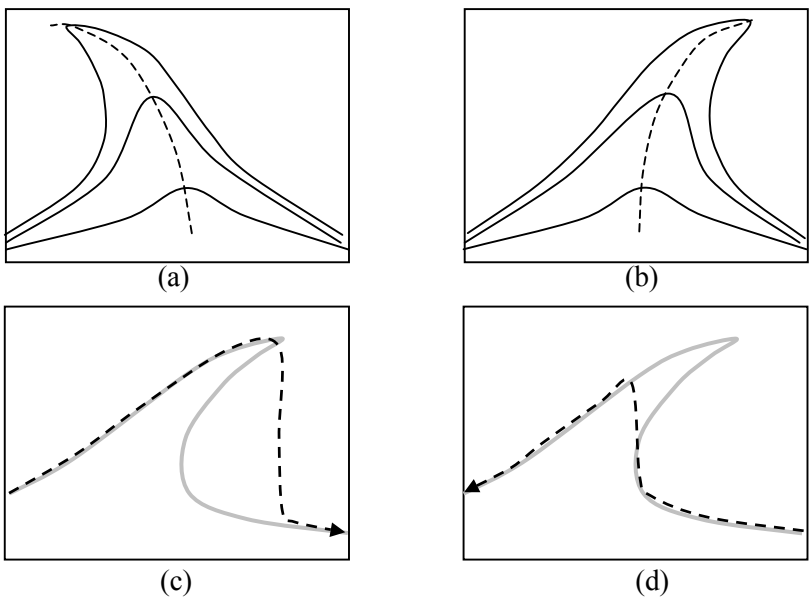


Figure 5.5: *Frequency-pulling effect and jump phenomenon of a Duffing oscillator. (a) Response curves of a softening spring system under various driving magnitudes. (b) Response curves of a hardening spring system. (c) Response curve of a hardening spring system under an increasing driving frequency. (d) Response curve of a hardening spring under a decreasing driving frequency.*

5.3.2 Harmonic balance

Considering both the cubic and quadric non-linearities, the membrane driven by a sinusoidal force has the equation of motion:

$$\ddot{x} + r\dot{x} + w_0^2 x + a x^2 + b x^3 = \frac{G_0}{m}\cos(wt + j_0)\,. \tag{5.25}$$

It can be rewritten as

$$\ddot{x} + r\dot{x} + w_0^2 x + a x^2 + b x^3 = \frac{G_0}{2m}(e^{jwt} + e^{-jwt})\,. \tag{5.26}$$

In the following calculation, the solution of this form is considered:

$$x = \sum_{n=-\infty}^{\infty} \tilde{B}_n e^{jnwt} \tag{5.27}$$

with $\tilde{B}_0 = 0$ and

$$\tilde{B}_{-n} = \tilde{B}_n^{\,*}\,. \tag{5.28}$$

Substituting Eq. (5.27) to Eq. (5.26) gives

$$\begin{aligned} &-w^2 \sum_{n=-\infty}^{\infty} n^2 \tilde{B}_n e^{jnwt} + jrw \sum_{n=-\infty}^{\infty} n\tilde{B}_n e^{jnwt} + w_0^2 \sum_{n=-\infty}^{\infty} \tilde{B}_n e^{jnwt} y \\ &+ a \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} \tilde{B}_n \tilde{B}_m e^{j(n+m)wt} + b \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \tilde{B}_n \tilde{B}_m \tilde{B}_l e^{j(n+m+l)wt} = \frac{G_0}{2m}(e^{-jwt} + e^{jwt}) \end{aligned} \tag{5.29}$$

Equating terms of various harmonics in the Fourier series yields a set of equations relating the series coefficients. In exact form, they are not amenable to simple calculation. This technique, sometimes called *harmonics balance* (see, e.g., Woodhouse 1995), is especially useful in small-amplitude vibrations. Therefore, $\xi(t)$ is assumed to be predominated by the fundamental and B_n decreases rapidly with n. In the following calculation, all B_n for $n > 3$ are neglected.

Collecting the terms of $e^{2j\omega t}$ in Eq. (5.29) gives

$$\begin{aligned} &(w_0^2 - 4w^2 + 2jrw)\tilde{B}_2 + a(\tilde{B}_1^2 + 2\tilde{B}_{-1}\tilde{B}_3) \\ &+ b(6\tilde{B}_1\tilde{B}_{-1}\tilde{B}_2 + 3\tilde{B}_2\tilde{B}_{-2}\tilde{B}_2 + 6\tilde{B}_3\tilde{B}_{-3}\tilde{B}_2 + 6\tilde{B}_{-2}\tilde{B}_1\tilde{B}_3) = 0 \end{aligned} \tag{5.30}$$

Assuming

$$\left|\tilde{B}_1\right| >> \left|\tilde{B}_2\right| \quad \text{and} \quad \left|\tilde{B}_1\right| >> \left|\tilde{B}_3\right|, \tag{5.31}$$

Eq. (5.30) becomes

$$(w_0^2 - 4w^2 + 2jrw)\tilde{B}_2 + a\tilde{B}_1^2 + 6b\left|\tilde{B}_1\right|^2 \tilde{B}_2 = 0\,. \tag{5.32}$$

$$\Rightarrow \widetilde{B}_2 = \frac{-\mathrm{a}\widetilde{B}_1^{\,2}}{\mathrm{w}_0^{\,2} - 4\mathrm{w}^2 + 2jr\mathrm{w} + 6\mathrm{b}\left|\widetilde{B}_1\right|^2} \tag{5.33}$$

Similarly, collecting the terms of $e^{3j\omega t}$ in Eq. (5.29) gives

$$\begin{aligned}&(\mathrm{w}_0^{\,2} - 9\mathrm{w}^2 + 3jr\mathrm{w})\widetilde{B}_3 + 2\mathrm{a}\widetilde{B}_1\widetilde{B}_2 \\ &+ \mathrm{b}(\widetilde{B}_1\widetilde{B}_1\widetilde{B}_1 + 6\widetilde{B}_1\widetilde{B}_{-1}\widetilde{B}_3 + 6\widetilde{B}_2\widetilde{B}_{-2}\widetilde{B}_3 + 3\widetilde{B}_3\widetilde{B}_{-3}\widetilde{B}_3 + 3\widetilde{B}_{-1}\widetilde{B}_2\widetilde{B}_2) = 0\end{aligned} \tag{5.34}$$

According to Eq. (5.31), Eq. (5.34) becomes

$$(\mathrm{w}_0^{\,2} - 9\mathrm{w}^2 + 3jr\mathrm{w})\widetilde{B}_3 + 2\mathrm{a}\widetilde{B}_1\widetilde{B}_2 + \mathrm{b}\widetilde{B}_1^{\,3} = 0. \tag{5.35}$$

5.3.3 Worman's theorem

Eqs. (5.33) and (5.35) are especially useful when the membrane's non-linear property is predominated by either the quadric term or the cubic term.

If $2\mathrm{a}\left|\widetilde{B}_2\right| << \mathrm{b}\left|\widetilde{B}_1\right|^2$, corresponding to the predominance of cubic non-linearity, Eq. (5.35) gives

$$\left|\widetilde{B}_3\right| = \frac{\mathrm{b}\left|\widetilde{B}_1\right|^3}{\sqrt{(\mathrm{w}_0^{\,2} - 9\mathrm{w}^2)^2 + (3r\mathrm{w})^2}}. \tag{5.36}$$

Conversely, if $2\mathrm{a}\left|\widetilde{B}_2\right| >> \mathrm{b}\left|\widetilde{B}_1\right|^2$, corresponding to the predominance of quadric non-linearity, Eq. (5.35) gives

$$\left|\widetilde{B}_3\right| = \frac{2\mathrm{a}\left|\widetilde{B}_1\right|\left|\widetilde{B}_2\right|}{\sqrt{(\mathrm{w}_0^{\,2} - 9\mathrm{w}^2)^2 + (3r\mathrm{w})^2}}. \tag{5.37}$$

If β is so small that

$$\left|\mathrm{w}_0^{\,2} - 4\mathrm{w}^2\right| >> 6\mathrm{b}\left|\widetilde{B}_1\right|^2, \tag{5.38}$$

Eq. (5.33) becomes

$$\left|\widetilde{B}_2\right| = \frac{\mathrm{a}\left|\widetilde{B}_1\right|^2}{\sqrt{(\mathrm{w}_0^{\,2} - 4\mathrm{w}^2)^2 + (2r\mathrm{w})^2}}. \tag{5.39}$$

Combining with Eq. (5.37) yields

$$\left|\widetilde{B}_2\right| \propto \left|\widetilde{B}_1\right|^2 \quad \text{and} \quad \left|\widetilde{B}_3\right| \propto \left|\widetilde{B}_1\right|^3. \tag{5.40}$$

These relations between the magnitudes of the lowest three harmonics of the membrane's displacement take the form of Worman's (1971) theorem when quadric non-linearity predominates its behavior. Worman's theorem states that when the vibration of the clarinet is small, the magnitude of the *n*th harmonic is proportional to the *n*th power of the magnitude of the fundamental:

$$\left|\widetilde{B}_n\right| \propto \left|\widetilde{B}_1\right|^n. \tag{5.41}$$

In practice this effect is brought into play by increasing the blowing pressure a little above the threshold to make the note louder. The content of higher harmonics grows more significantly than lower harmonics. Although the sounding mechanism underlying the clarinet is very different from that of the *dizi*, a similar law is shown to govern their harmonic generation in small oscillations near the threshold.

If cubic non-linearity predominates the membrane's behavior, Eq. (5.36) also takes the form of Worman's theorem. According to Eq. (5.33), the magnitude of the second harmonic is also approximately proportional to the square of the fundamental magnitude unless the driving frequency is very close to the $f_m/4$. However, the second harmonics is much weaker than the third harmonics because the former is proportional to the coefficient of quadric non-linearity, which is assumed to be small.

In summary, the predominance of cubic non-linearity of the membrane has two signatures. First, the magnitude of the third harmonic is proportional to the third power of the fundamental magnitude,. Second, the second harmonic is weaker than the third harmonic. On the other hand, the predominance of the quadric non-linearity is characterized by Eq. (5.40), which takes the form of Worman's theorem, and the second harmonic is stronger than the third harmonic in oscillations near the threshold. If there is no such simple relation between the lowest three harmonics, neither quadric non-linearity nor cubic non-linearity predominates.

5.4 Experimental study

5.4.1 Experimental setup

The purpose of this experiment is to demonstrate the existence of the *dizi* membrane's non-linearity. As wrinkling in the membrane plays a crucial role in producing a "correct" *dizi* timbre,

this experiment was also dedicated to the influences of wrinkling on its non-linearity. The predominance of the quadric or cubic non-linearity was judged by comparing the amplitude of the lowest three harmonics. The response curves were measured to determine the sign of the cubic non-linearity and to estimate the resonant frequency of the membrane.

Fig. 5.6 shows the general features of the experimental arrangement, which was designed to apply sufficiently intense sinusoidal signals to the membrane. The loudspeaker was a medium range dome type, 18 cm in diameter. It was mounted with a horn made of pottery, 17 cm in tallness and 6 mm in thickness. The horn's thickness was expected to reduce the wall vibration, which would likely evoked by loud sounds in the cavity and lead to errors in experiments. The inner wall of the horn was conically shaped. It is desirable to avoid strong resonances of the cavity. Strong resonances would introduce an unnecessary strong frequency dependence of the acoustic pressure at the mouth. A plate with a hole sealed by a membrane covered the opening of the horn and a microphone was placed above the membrane with a distance of approximately 1 cm.

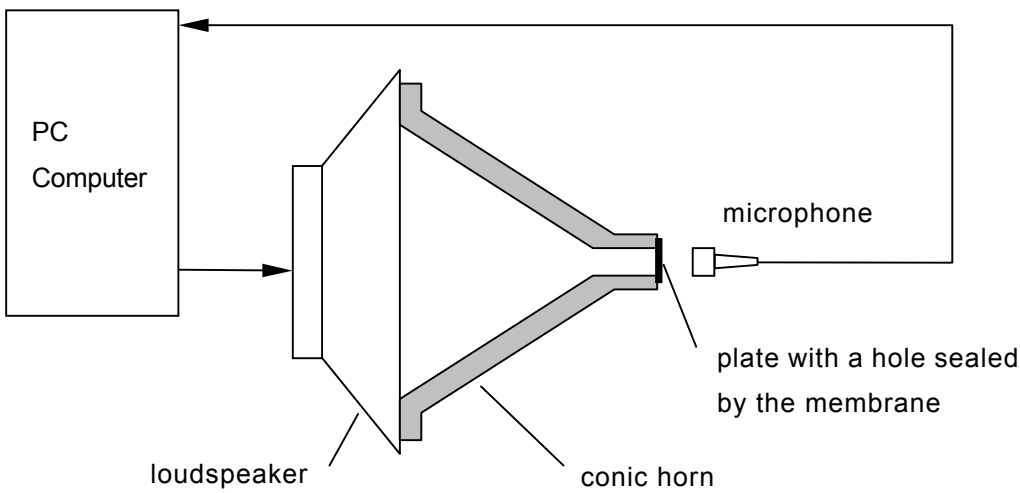


Figure 5.6: *Experimental setup.*

5.4.2 Materials

Sinusoidal signals were synthesized by a PC computer. First, two signals with a constant amplitude and the frequency monotonically increasing/decreasing from 2 kHz to 3 kHz were generated. They were calibrated by the acoustic pressure level measured at the mouth of the conic horn without the plate. After calibration, I obtained the input signals with approximately constant acoustic pressure for driving the membrane placed at the opening of the horn. Although this calibration was rather crude, the resonance of the membrane was expected to be strong enough to mask the errors due to amplitude fluctuations of the input signal while its frequency swept up/down.

Second, a signal with a constant frequency 1350 Hz and monotonically increasing amplitude was generated.

5.4.3 Experimental procedure

Two measurements were performed on four states of the membrane. In the first measurement, the membrane was driven by the calibrated signals with increasing/decreasing frequency. The magnitude of the input signal was modified four times. The sounds radiated by the membrane were recorded by the microphone above it. The magnitude of the fundamental was plotted along the frequency and the four response curves ware compared. Owing to the limitation of equipments, the sounds radiated by the membrane were recorded in an arbitrary unit.

In the second measurement, the membrane was driven by a signal with a constant frequency 1350 Hz and monotonically increasing amplitude. The sounds radiated by the membrane were decomposed into Fourier components.

These two measurements were taken for four states of the membrane. The membrane was first sealed on a hole with a radius of 4 mm in a piece of cylindrically shaped plate. The cylinder has a radius of 30 mm. Then the same measurements were taken for the membrane sealed on a hole in a flat plate. This procedure was carried out for a wrinkled membrane and an unwrinkled membrane.

Finally, it should be noted that the coefficients of the Fourier components of the sound radiated by the membrane are approximately proportional to the membrane's acceleration:

$$A_n \propto n^2 \left| \widetilde{B}_n \right| . \tag{5.42}$$

This approximation is based on the assumption that the vibrating membrane is baffled in an infinite plane. Referring to Eq. (7.28) in [Fletcher and Rossing 1991] for such a sound source, the pressure dp produced by a small element of area dS at r' in a plane with velocity $u(r')$ is

$$dp(r) = \frac{j\omega\rho_0}{2\pi r} e^{-jk|r-r'|} u(r') dS . \tag{5.43}$$

Assuming that the velocity $u(r')$ is uniform at every points of the membrane (piston approximation) and the wavelength is much larger than the distance between the membrane and the microphone ($kr \sim 0$), the integral over the membrane gives

$$p(r) = \frac{j\omega\rho_0}{2\pi r} u S_m . \tag{5.44}$$

As it is in proportion to ωu for a sinusoidally vibrating membrane, the acoustic pressure recorded above the membrane is proportional to its acceleration.

5.5 Results and discussion

5.5.1 Magnitudes of the lowest three harmonics

Figs. 5.7 and 5.8 give the response curves (left) for signals with sweeping up/down frequency and the relations between the magnitudes of the lowest three harmonics (right) of the sounds radiated from the membrane. The behavior of the membrane sealed on a hole in the wall of a cylindrically shaped plate showed a predominance of quadratic non-linearity when it was wrinkled. As can be seen from Fig. 5.7(a-2), the amplitude of the second (third) harmonic was proportional to the second (third) power of the amplitude of the fundamental, and the second harmonic was stronger than the third harmonic. When the intensity of the driving signal became larger, however, the amplitude of the second and third harmonics deviated from Worman's theorem. This can be explained by the Taylor expansion of the membrane's stiffness; when its vibration amplitude is large enough, both quadric and cubic non-linearities contribute to the membrane's dynamic behavior and there are no simple relations between the amplitudes of the lowest three harmonics.

When the membrane sealed on the cylindrically shaped plate was unwrinkled, the third harmonic was always stronger than the second harmonic. When the intensity of the driving signal was large enough, the amplitude of the third harmonics was proportional to the third power of the amplitude of the fundamental [Fig. 5.7(b-2)]. This implies a predominance of cubic non-linearity in this range. When the intensity of the driving was small, however, the membrane's harmonic generation did not obey Worman's theorem. This can also be explained by the Taylor expansion of the membrane's stiffness; when its vibration amplitude is small, the quadric term is non-negligible even when the coefficient of the quadric term is small. The cubic term does not yet predominate in small-amplitude vibrations and both quadric and cubic non-linearities contribute to the membrane's dynamic behavior. In contrast to the wrinkled membrane, the quadric non-linearity never predominated the behavior of the unwrinkled membrane.

The behavior of the membrane sealed on a flat plate was quite similar to that sealed on a curved plate. The wrinkled membrane showed a predominance of quadric non-linearity [Fig. 5.8(a-2)], whereas the unwrinkled membrane showed a predominance of cubic non-linearity [Fig. 5.8(b-2)]. Comparing Fig. 5.7(b-2) to Fig. 5.8(b-2) yields that the unwrinkled membrane sealed on a flat plate has a predominance of cubic non-linearity even when the intensity of the driving signal is small. This implies that the unwrinkled membrane sealed on a hole in a cylinder has a larger value of the quadric non-linearity. This supports that the membrane's global curvatures break the symmetry between its upward and downward deformations (section 5.2.1.2).

It should be noted that Worman's theorem holds only for extreme values of the criterion $2\mathrm{a}\left|\widetilde{B}_2\right|\Big/\mathrm{b}\left|\widetilde{B}_1\right|^2$ mentioned in section 5.3.2. When $2\mathrm{a}\left|\widetilde{B}_2\right|\Big/\mathrm{b}\left|\widetilde{B}_1\right|^2 >>1$ the predominance of quadric non-linearity leads to $\left|\widetilde{B}_2\right| \propto \left|\widetilde{B}_1\right|^2$ and $\left|\widetilde{B}_3\right| \propto \left|\widetilde{B}_1\right|^3$; when $2\mathrm{a}\left|\widetilde{B}_2\right|\Big/\mathrm{b}\left|\widetilde{B}_1\right|^2 <<1$ the predominance

of cubic non-linearity leads to $\left|\widetilde{B}_3\right| \propto \left|\widetilde{B}_1\right|^3$. In the present study the values of the amplitude of membrane's vibration, quadric non-linearity and cubic non-linearity were not obtained. The use of Worman's theorem remains to be justified by quantitative experiments in the future.

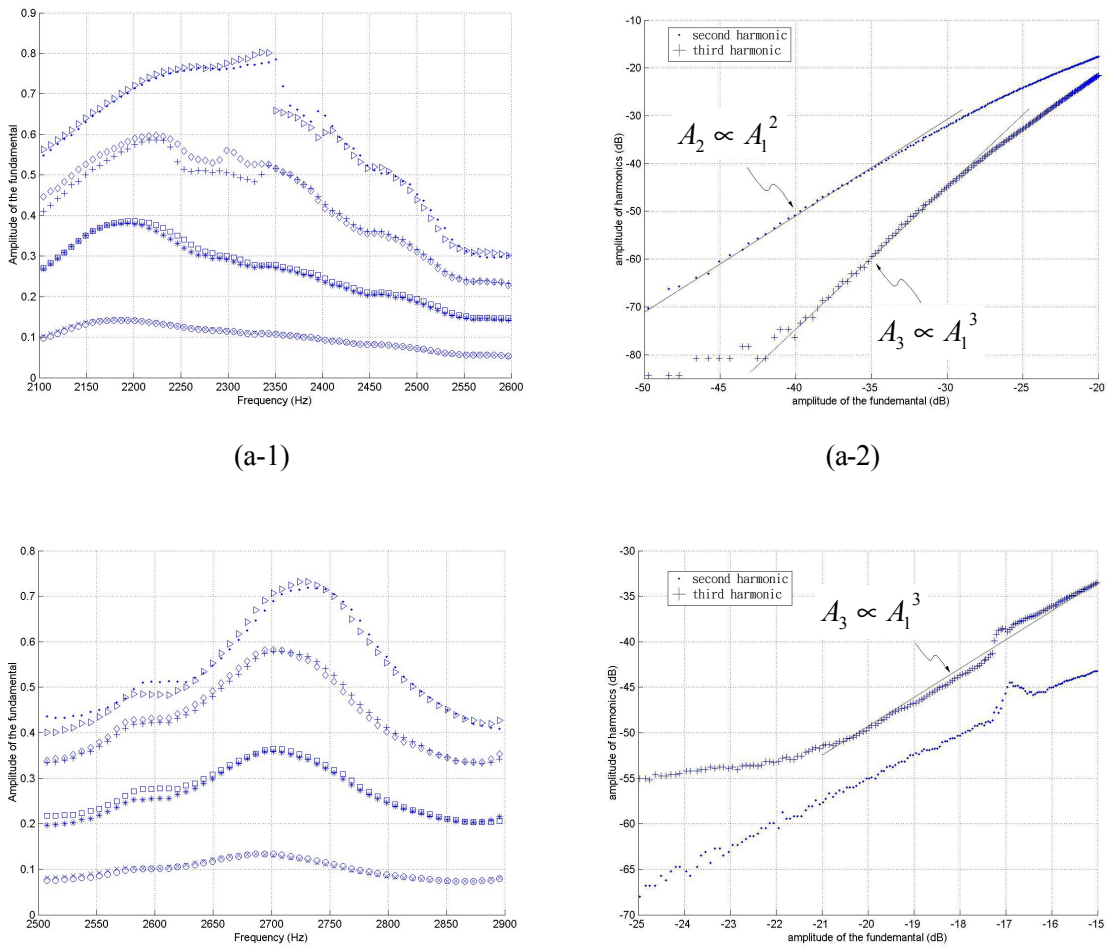


Figure 5.7: *Response curves (left) and relations between the amplitudes of lowest three harmonics (right) of the sounds radiated from a membrane sealed on a hole in the wall of a cylindrical plate. In the response curves the plots marked with '+ ∗ × ·' are measurements for a increasing driving frequency, while those marked with '△ ◊ □ ○' are measurements for a decreasing driving frequency. (a) Wrinkled membrane. (b) Unwrinkled membrane.*

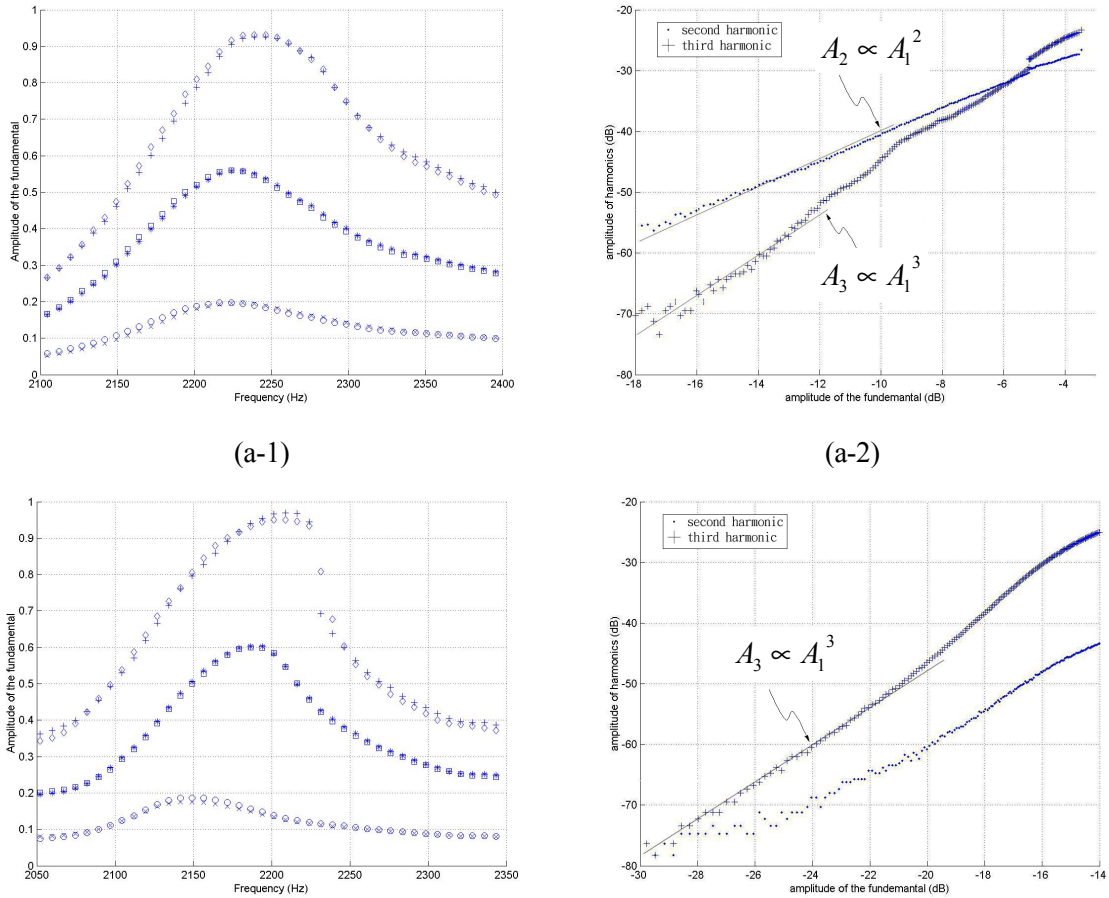


Figure 5.8: *Response curves (left) and relations between the amplitudes of lowest three harmonics (right) of the sounds radiated from a membrane sealed on a hole in the wall of a flat plate. In the response curves the plots marked with '∗ × · ' are measurements for a increasing driving frequency, while those marked with '◊ □ ○' are measurements for a decreasing driving frequency. (a) Wrinkled membrane. (b) Unwrinkled membrane.*

5.5.2 Response curves

The response curves of the membrane show the frequency-pulling effect characteristic of a hardening spring regardless of wrinkling [Fig. 5.7(a-1)(b-1), Fig. 5.8(a-1)(b-1)]. The maximum of the response curves tends to increase as the amplitude of the driving signal increased. For the membrane sealed on a hole in a flat plate, wrinkling largely reduced the frequency-pulling effect [Fig. 5.8(a-1)]. This supports that a wrinkled membrane has a lower value of cubic non-linearity than an unwrinkled membrane for the same resonant frequency.

For the membrane sealed on a hole in the wall of a cylinder, however, the response curves are complicated by the involvement of the quadric non-linearity due to the membrane's global curvatures. This might explain the zigzags in the response curve for the wrinkled membrane [Fig. 5.7(b-1)]. In

such large-amplitude vibrations, the geometry constrain of inextensibility may be not satisfied and the non-linearity contribute to the dynamic behavior of the membrane is complicated. The irregular response curve of the wrinkled membrane sealed on a hole in the wall of a cylinder requires further investigations.

Fig. 5.8(b-1) indicates that the membrane's response changed rapidly at 2230 Hz for a strong driving signal. When I imposed a lower initial strain to this unwrinkled membrane, its cubic non-linearity became so large that its response curve showed a jump phenomenon (Fig. 5.9). Note that the signals used in the present experiment were much weaker than the acoustic waves in the *dizi* pipe in a typical playing range. Therefore, the jump phenomenon is likely to occur in a normal play condition when the *dizi* membrane is unwrinkled and slack.

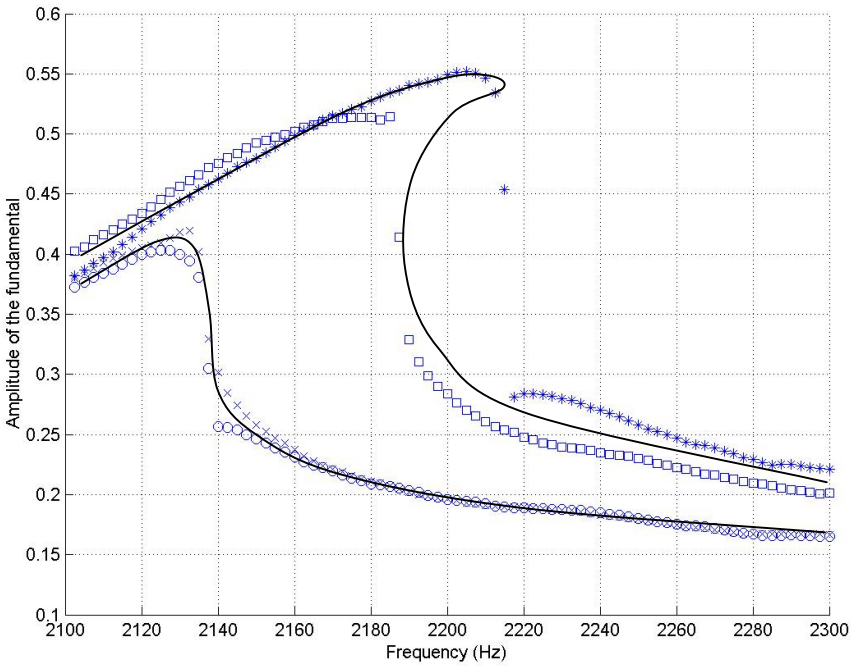


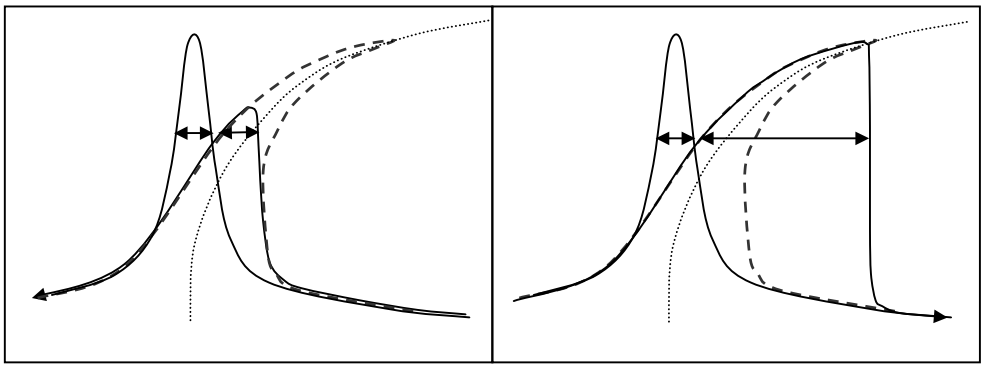
Figure 5.9: *Jump phenomenon in the response curve of a slack, unwrinkled membrane under the driving frequency sweeping up* (∗, ×) *and down* (□, ○).

5.6 Jump phenomenon, wrinkling and performance practice

From a physical point of view, one of the most interesting properties of the Duffing oscillator is the jump phenomenon. In the realm of music acoustics, a string has been proved to have a hardening spring characteristic. A jump phenomenon of an electromagnetically driven string was reported by Valette (1993).

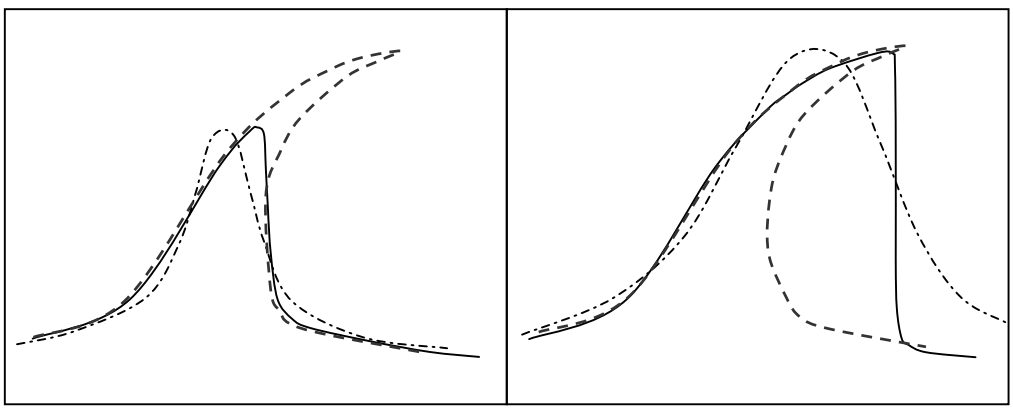
The jump phenomenon of the *dizi* membrane occurs when it is slack and unwrinkled so that its cubic non-linearity is large (Fig. 5.9). This phenomenon can be related to playability reductions for

the tones in the second and third registers. For the same fingering, the *dizi* either does not sound, or sounds raucously. Additionally, these tones have a strong tendency to jump to higher play modes, where the membrane is located near a pressure node of the fundamental. These phenomena can be explained in terms of the response curve of a Duffing oscillator.



(a) (b)

Figure 5.10: *Q-factor reductions of the membrane's resonance due to the jump phenomenon. (a) Small jump at a lower frequency. The Q-factor of the resonance is slightly smaller than that of a linear membrane. (b) Large jump at a higher frequency. The Q-factor of the resonance is significantly smaller than that of a linear membrane.*



(a) (b)

Figure 5.11: *Approximations of the membrane's resonance with a jump. (a) Small jump at a lower frequency. The resonance curve is approximated as that of a damped harmonic oscillator with a lower resonant frequency and a lower damping coefficient. (a) Large jump at a higher frequency. The resonance curve is approximated as that of a damped harmonic oscillator with a higher resonant frequency and a higher damping coefficient.*

Owing to the jump phenomenon of the membrane, its resonance has different Q-factors corresponding to a decreasing or increasing drive frequency. When the drive frequency sweeps down, the resonance of a non-linear membrane has a Q-factor slightly lower than that of a linear membrane [Fig. 5.10(a)]. On the other hand, when the drive frequency sweeps up, the resonance of a non-linear membrane has a Q-factor much lower than that of a linear membrane [Fig. 5.10(b)]. As the Q-factor represents the damping coefficient of the membrane, a non-linear membrane can have different values of the effective damping coefficient. This considerably affects the input impedance curve of the *dizi* through the pipe-membrane coupling.

With a slack, unwrinkled membrane, the *dizi* is difficult to play in the second and third registers, for which the membrane is located near a pressure antinode of the fundamental wave in the pipe. In the second register the *dizi* tends to jump from the first play mode to the second play mode; in the third register, the *dizi* tends to jump from the second play mode to the third play mode. This unstable behavior due to a slack, unwrinkled membrane can be explained in terms of the input impedance of the *dizi*.

The exact expression of the response curve of a non-linear membrane can be obtained by solving Eq. (5.24). For the first approximation, instead, I use the form of a damped harmonic oscillator for impedance calculation. The jumps at low and high frequencies are associated with different resonant frequencies and damping coefficients, as shown in Fig. 5.11. The effective damping coefficient corresponding to a large jump [Fig. 5.11(b)] is larger than that corresponding to a small jump [Fig. 5.11(a)]. Given the membrane's resonant frequency, damping coefficient and mass, the input impedance of the *dizi* can be calculated (see section 4.2.2). In order to simulate the jump phenomenon of a *dizi* membrane, these parameters are chosen to mimic the membrane's response curve shown in Fig. 5.9. The resonant frequency of the large jump is estimated as 100 Hz higher than that of a small jump, and the damping coefficient corresponding to a large jump is estimated as three times as that corresponding to a small jump. This estimation is based on the membrane's large-amplitude vibrations in a typical playing range, in which the driving force is much larger than the signals used in the previous experiments. Consequently, the two jumps are separated in a distance larger than that showed in Fig. 5.9(a). This implies a large value of the effective damping coefficient corresponding to the jump at the higher frequency.

The results of calculated input admittance curves of a *dizi* with a slack, unwrinkled membrane are given in Fig. 5.12. Fig. 5.12(a-1) shows the admittance curve of the *dizi* under the fingering ●○○○○○ with a jump at a lower frequency [Fig. 5.11(a)], while Fig. 5.12(a-2) shows the admittance curve with a jump at a higher frequency [Fig. 5.11(b)]. Similarly, Figs. 5.12(b-1) and (b-2) show the two admittance curves of the *dizi* under the fingering ●●●●○○.

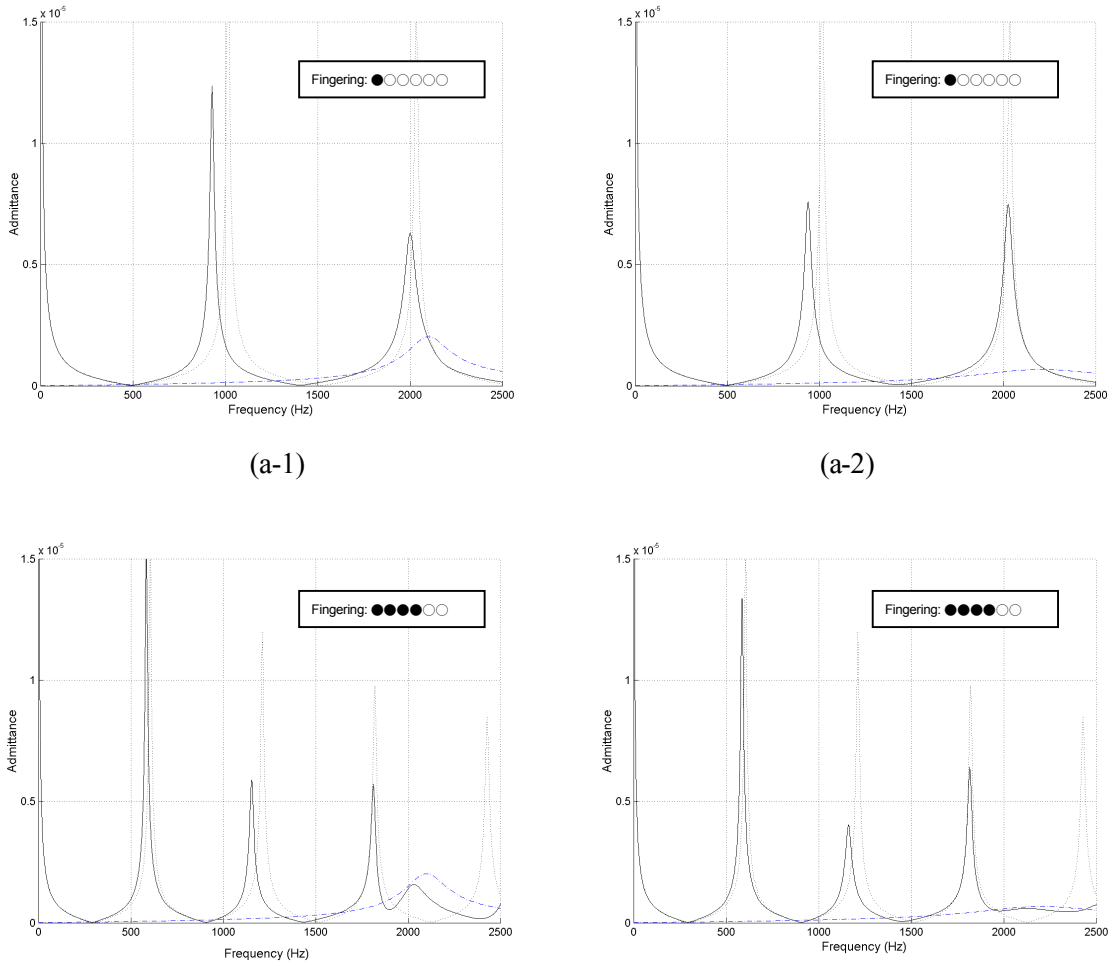


Figure 5.12: *Examples of the theoretical admittance curves of the dizi and its membrane. Comparison between the pipe without membrane vibration (- - -), the pipe with membrane vibration (—) and the membrane's acoustic admittance curve (— -). Mass of the membrane m = 1.4×10^{-6} kg. Radius of the pipe r_0 = 8 mm. Radius of the membrane hole a = 4.5 mm. (a) Distance between the membrane and the mouth hole L_1 = 90 mm and distance between the membrane and the first open finger hole L_1 = 80 mm. These two parameters mimic the fingering* ●○○○○○ *of the qudi in C. (b) L_1 = 90 mm and L_1 = 195 mm. They mimic the fingering* ●●●●○○ *of the qudi in C. (1) Damping coefficient R = 2×10^{-3} sec^{-1}kg. Resonant frequency of the membrane f_m = 2.1 kHz. These parameters mimic the jump phenomenon at the lower frequency [Fig. 5.11(a)]. (2) R = 6×10^{-3} sec^{-1}kg. f_m = 2.2 kHz. These parameters mimic the jump phenomenon at the higher frequency [Fig. 5.11(b)]. Comparing the heights of the peaks in (a) and (b), we note reductions in playability of the first pipe mode under the fingering* ●○○○○○ *for a large jump. Similarly, the second mode under the fingering* ●●●●○○ *is more difficult to be excited than the third mode for a large jump.*

Figs. 5.12(a-1) and (a-2) show how the jump phenomenon of a slack, unwrinkled membrane reduces the playability of a *dizi* tone in the second register (fingering ●○○○○○), where the membrane is located at a pressure antinode of the fundamental wave. The *dizi* note in the first play mode (second register) is represented by the peak at 0.95 kHz, while the *dizi* tone in the second play mode (fourth register) is represented by the peak at 2 kHz. If the jump of the membrane's response curve is at the lower frequency, the resonance at 0.95 kHz is more easily to excited, as its admittance is higher and that at 2 kHz [Fig. 5.12(a-1)]. But if the jump of the membrane's response curve is at the higher frequency, a reduction in admittance of the resonance at 0.95 kHz will lead to a tendency to excite the resonance at 2 kHz, as their admittances are approximately the same [Fig. 5.12(a-2)].

For the *dizi* tones in the third register (second play mode), their playability reductions due to the jump phenomenon are even more significant. Figs. 5.12(b-1) and (b-2) compare the resonances for the second and third modes under the fingering ●●●●○○. For a jump of the membrane's response curve at the lower frequency, the second mode is easier to be excited than the third mode [Fig. 5.12(b-1)]. On the contrary, for a jump at the higher frequency, the third mode is easier to be excited, while the second mode cannot be excited because of a low admittance of the resonance [Fig. 5.12(b-2)]. This explains the strong tendency of the *dizi* to jump from the second mode to third mode when the membrane is unwrinkled and slack.

It is important to distinguish the playability reductions of the tones in the second and third registers due to the membrane's jump phenomenon from those of the tones in the fifth register due to a low resonant frequency of the membrane (see chapter 4). The *dizi* tones in the second and third registers have fundamental frequencies much lower than the membrane's resonant frequency. Consequently, the admittance reductions for these tones are unlikely to result from a low resonant frequency of the membrane. Instead, these admittance reductions stem from a high value of the effective damping coefficient associated with a jump phenomenon.

The theoretical discussion cited above provides an explanation for the discontinuous behavior of the *dizi* with a slack, unwrinkled membrane. The multi-valued response curve of a slack, unwrinkled membrane implies that for a driving frequency in this range, the membrane has a stable state and a meta-stable state. Its behavior is sensitive to its history, thus leading to a hysteresis of the *dizi*. Wrinkling largely stabilizes *dizi* sound production and avoids the jump phenomenon through reduction in cubic non-linearity.

The prediction of different admittance curves of the *dizi* for sweeping up/down driving signals shown in Fig. 5.12 should be tested in future experiments. Such impedance measurements on the *dizi* require a technique using intense sweeping frequencies as input signals.

5.7 Conclusions

The existence of membrane's non-linearity of the lowest two orders was demonstrated by driving it with sinusoidal acoustic signals with an increasing intensity or a sweeping frequency. The

predominance of cubic or quadric non-linearity was observed under different conditions of the membrane with a resonant frequency in the range of 2–3 kHz. The behavior of an unwrinkled membrane is predominated by cubic non-linearity, whereas a wrinkled membrane showed a predominance of quadric non-linearity in an intensity range of the input signal. This supports that wrinkling reduces the membrane's cubic non-linearity. When the membrane is slack and unwrinkled, its non-linearity is large enough to result in a jump phenomenon in its response curve.

Impedance calculations explain the low playability of the *dizi* tones in the second and third registers in terms of the jump phenomenon. The membrane with a large jump in its response curve is approximated as a linear oscillator with a large damping coefficient. This damping reduces the admittances of the bore resonances supporting notes in the second and third registers, so the *dizi* tends to jump to higher resonances with higher admittances.

References

Cerda, E., and Mahadevan, L. (2003). Geometry and physics of wrinkling. Physical Review Letters 90, 074302.

Chobotov, V.A., and Binder, R.C. (1964). Nonlinear response of a circular membrane to sinusoidal acoustic excitation, Journal of the Acoustical Society of America 36, 59-71.

Hagedorn, P. (1995). Mechanical oscillations. In: Mechanics of Musical Instruments (edited by A. Hirschberg, J. Kergomard and G. Weinreich), Springer-Verlag, 7-78.

Landau, L., and Lifshitz, E.M. (1986). Theory of Elasticity. Pergamon.

Thrasher, A. (2001). "*Di*" in The New Grove Dictionary (edited by S. Sadie), Macmillan, Vol. 7, 277-279.

Valette, C. (1995). The mechanics of vibrating strings. In: Mechanics of Musical Instruments (edited by A. Hirschberg, J. Kergomard and G. Weinreich), Springer-Verlag, 115-184.

Worman, W.E. (1971). Self-Sustained Nolinear Oscillations of Medium Amplitude in Clarinet-like Systems. PhD thesis, Case Western Reserve University.

Woodhouse J. (1995). Self-sustained musical oscillators. In: Mechanics of Musical Instruments (edited by A. Hirschberg, J. Kergomard and G. Weinreich), Springer-Verlag, 185-228.

Chapter 6.
Spectral features of *dizi* tones and the membrane's cubic non-linearity

Abstract

The *dizi* membrane is modeled as a Duffing oscillator driven by the acoustic waves in the tube. This model aims to explain three major spectral features of *dizi* tones: (1) formants, (2) the predominance of odd-numbered harmonics, and (3) the subharmonics at high frequencies. It was tested by two kinds of quasi-sinusoidal tones driving the membrane: tones generated by external excitation or by blowing the instrument. For external excitation, the phase plots of the membrane show interlocking spirals, which are predicted by the quasi-sinusoidally driven Duffing oscillator model. For normal *dizi* tones, the agreement is restricted to frequencies < 10 kHz. In principle, the first formant, the odd-numbered harmonics predominance, and subharmonics at high frequencies are successfully explained by the membrane's cubic non-linearity.

Rich harmonics above 10 kHz radiated by the membrane, not predicted by our model, may be associated with the jet's sensitivity to the harmonics in the range 3–5 kHz generated by the membrane. If the jet amplifies these harmonics in the pipe, they may in turn drive the membrane to radiate harmonics above 10 kHz. Two mechanisms relevant to the jet's sensitivity to these harmonics are proposed: the moving separating points of the jet, and its varicose oscillations due to the pressure fluctuations at the flue exit. While these two effects are difficult to demonstrate experimentally, the second formant frequency can be explained by the filtering effect of the lowest transverse pipe mode with the resonance estimated as $c/2D$ (D is the pipe's diameter).

6.1 Introduction

In chapter 5, the existence of cubic and quadric non-linearities of the *dizi* membrane was demonstrated. The goal of this chapter is to explain the characteristic *dizi* timbre in terms of the membrane's cubic non-linearity. A Duffing oscillator model of membrane is proposed to account for

three major spectral features of *dizi* tones: formants, the odd-numbered harmonics predominance, and subharmonics at high frequencies.

Before taking the membrane into consideration, it is helpful to review the physical modeling of flue instruments. The geometry of a typical recorder-like instrument is illustrated in Fig. 6.1. A steady pressure supply in the foot of the pipe generates an air flow blowing through a narrow windway or flue. At the flue exit, a free jet is formed by flow separation. This jet flows across the mouth of the instrument and is directed toward a sharp edge called the *labium*. The jet is submitted to the transverse acoustic flow due to the air column's oscillation in the pipe. The jet is sensitive to these perturbations because of its intrinsic instability. As the jet travels toward the labium, these perturbations are amplified. This results in a flipping of the jet on each side of the labium at the same frequency of the acoustic field. This jet motion sustains the air column's oscillation.

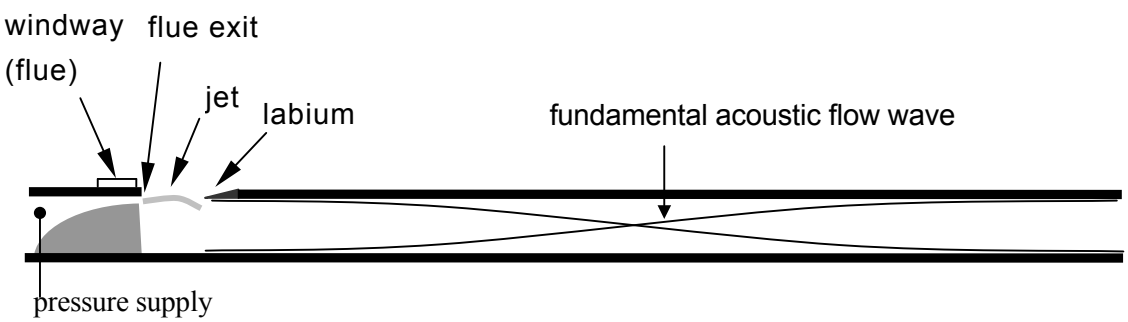


Figure 6.1: *Geometry of a recorder-like flue instrument.*

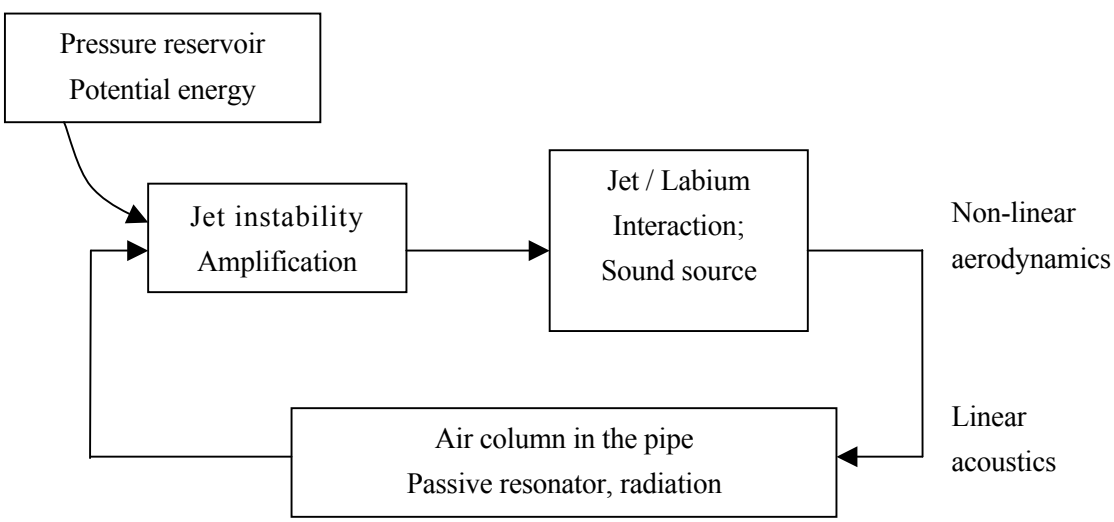


Figure 6.2: *A simplified description of a flue instrument in a lumped model.*

The functioning of recorder-like instruments is usually represented by a feedback loop composed of several elements: the jet instability is the amplifier; the pipe is an acoustic linear filter; and the jet-labium interaction is the sound source. Fig. 6.2 presents a simplified description of the feedback loop in a flue instrument (Fabre *et al.* 2000). The elements of this loop could first be isolated from the system for specific study, and then lumped together. This is a basic assumption of lumped models of flue instruments.

The role of the membrane in the *dizi* is comparable with the violin body, because it serves as the secondary resonator and the two-dimensional radiator of the instrument. Similarities between the *dizi* and the violin can be found in their elements. When a wind instrument or a bowed string instrument is played, most energy is preserved as standing waves in the linear, one-dimensional resonator: the air column in the *dizi*, and the strings of the violin. The oscillation of the one-dimensional resonator is sustained by the source: aerodynamics at the mouth hole of the *dizi*, and the Helmholtz motion of the violin strings. The source is clearly separated from the two-dimensional radiator: the membrane of the *dizi*, and the body of the violin. The one-dimensional resonator allows the source and the radiator to couple together, establishing a feedback loop. Only a limited amount of the energy preserved in the one-dimensional resonator is transferred to the two-dimensional radiator. The above analogue is summarized in Tab. 6.1.

	Violin	***Dizi***
Energy supply	Quasi-steady bow velocity	Quasi-steady jet velocity
Source	Bow-string interaction	Aerodynamics at the mouth hole
1-dimensional resonator	Strings	Air column
2-dimensional resonator and radiator	Soundboard	Membrane

Table 6.1: *Functions of mechanical elements of the dizi and the violin.*

Because the one-dimensional resonator of a musical instrument always has poor radiation efficiency, the two-dimensional resonator coupled to it always serves as the main radiator. In a membrane-less flute, the lumped air at the mouth hole and open finger-holes serves as the major radiators. However, there is no obvious separation between radiator, source and resonator at the mouth hole. This makes the physics of flue instruments much more intricate than that of the violin, whose body can be regarded as an independent element taking care of radiation. This basic difficulty of lumped models of flue instruments has been pointed out in a recent review (Fabre *et al*. 2000):

> It is a paradox that a flue instrument can be analyzed as a lumped system in a similar way to a violin, since there is no obvious separation between source, resonator and radiator. The key of this enigma is that the various flow phenomena involved correspond to different length and time scales, for which different approximations to the Navier-Stokes equations apply. But this is only true if we limit our discussion to low frequencies, corresponding to the fundamental and its first harmonics. Under such hypothesis, the sound production is limited to a region with a size of the order of the mouth, which is typically a fraction (10^{-2}) of the acoustic wavelength of the standing waves in the resonator. Furthermore, radiation at low frequencies is almost negligible. The radiation acoustic field is dominated by the lowest frequency components.[...]
>
> It is a striking aspect of the literature that most of the models are based on an intuitive approach in which approximations are not clearly defined. The discussion in the frame of lumped models often focuses on second order terms from a model which is at most accurate in the first order approximation.

The mechanics of flue instruments is governed by a set of non-linear equations: the Navier-Stokes equations. As a general exact solution of the Navier-Stokes equations is not available, solutions are built up out of approximations. These approximations are only valid in a restricted region of the flow. A global solution is obtained by "gluing" these local solutions together. At the mouth hole, the resonator, source and the radiation field are dominated by different frequency ranges. In order to describe various flow phenomena involved at the mouth hole, different approximations corresponding to different length and time scales should be considered with caution. This makes physical study on flue instruments rather challenging. For instance, only very limited results about the harmonic generation of flue instruments are available (Fletcher and Douglas 1980, Verge *et al.* 1997, Dequand 2000).

A lumped model of flue instruments is dramatically changed by the presence of a membrane. In a membrane-flute, the membrane is clearly separated from the source and the one-dimensional resonator. Although the membrane introduces complications to this system, this separation makes some simple approximations possible. For example, the harmonics generated at the mouth hole become ignorable in comparison to those generated by the non-linear membrane. Therefore, the difficult problem of harmonic generation can be accounted for by describing the motion of a forced non-linear oscillator, the *dizi* membrane. The membrane's stiffness is inherently non-linear even for soft *dizi* tones. This non-linear oscillator, able to transfer energy from lower frequencies to higher frequencies, serves as the main radiator for high frequencies.

The present research on non-linear mechanisms of the *dizi* is split into two parts. In the first part the membrane's behavior under a quasi-sinusoidal force is studied. This part begins by distinguishing between the radiation fields near the mouth hole and the membrane in section 6.2. Then, the membrane is modeled as a Duffing oscillator driven by a quasi-sinusoidal force for the *dizi* tones in the second and third registers. Section 6.3 addresses the experimental and numerical methods. The results are compared and discussed in section 6.4. A general explanation of spectral features of *dizi*

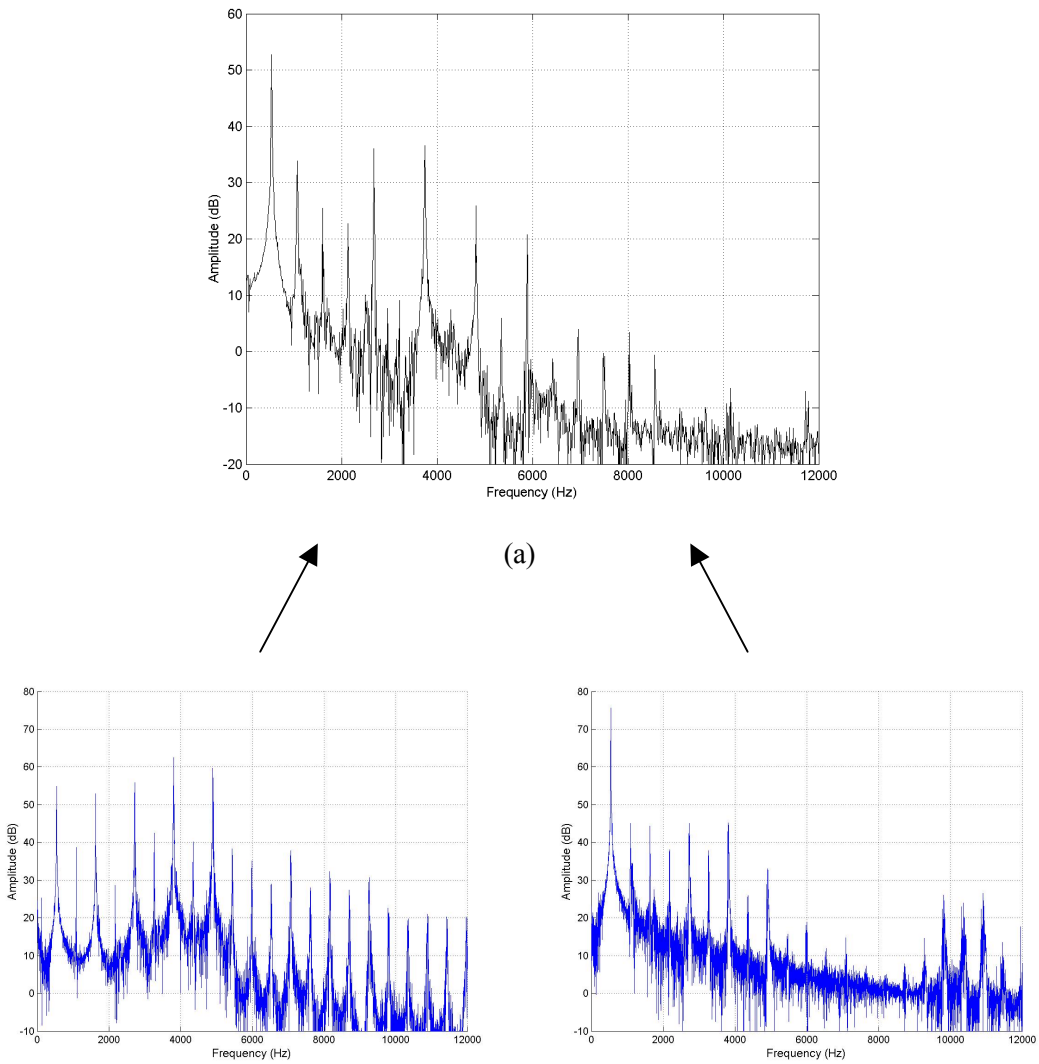
tones in five registers is provided in section 6.5. In the second part, the aerodynamics at the mouth hole is taken into account. Section 6.6 begins with a review of the drive mechanisms of flue instruments. Possible interactions between the membrane and the jet are discussed. The second formant of *dizi* tones is related to the filtering effect of pipe's transverse modes.

6.2 Spectral features of *dizi* tones

6.2.1 Tones radiated from the membrane and the mouth hole

In the experimental study described in chapter 5, acoustic signals were applied to the *dizi* membrane to demonstrate its lower-order non-linearity. These signals generated by a speaker were much weaker than the acoustic pressure waves in the pipe of a *dizi* during a performance. In a typical playing range, the acoustic pressure in the pipe is of the order of 10^2 Pa, which are always dominated by the fundamental wave. The period of its vibration is the same as the driving force, but the waveform is distorted and rich harmonics are generated. This implies that the radiation acoustic field near the membrane is predominated by high frequencies, while the radiation acoustic field near the mouth hole is predominated by the fundamental. The *dizi* tone reaching listeners is a combination of these radiations.

Fig. 6.1 compares three *dizi* tones played under approximately the same condition (fingering and embouchure), but recorded at different positions. Fig. 6.1(a) presents the spectrum of a *dizi* tone in the second register recorded in the front of the instrument. It has a dominant fundamental and a formant centered at 4 kHz. Amplitude differences between odd-numbered and even-numbered harmonics characteristic of this register appear only around the first formant. This spectrum can easily be explained in terms of the spectra of the two tones recorded near the mouth hole and the membrane. The tone radiated by the membrane shows significant amplitude differences between odd-numbered and even-numbered harmonics around and below the first formant [Fig. 6.1(b)]. At low frequencies, however, this characteristics is masked by the tone radiated from the mouth hole [Fig. 6.1(c)], whose lowest five harmonics do not show significant amplitude differences between odd-numbered and even-numbered harmonics. In addition, it is noteworthy that this tone has a formant centered at 10 kHz, where turbulence noises also accumulate.



(a)

(b) (c)

Figure 6.3: *Spectra of tones produced by a large dizi in G with the fingering* ○○○○○○○. *(a) Tone recorded 30 cm in front of the player. (b) Tone recorded 1 cm above the membrane. (c) Tone recorded 2 cm above the mouth hole.*

6.2.2 *Dizi* tones in the second and third registers

The following study accounts for the *dizi* tones in the second and third registers. The choice of these two registers was aimed to explain the three spectral characteristics of the *dizi* tones: (1) the two formants, (2) the predominance of odd-numbered harmonics, and (3) subharmonics at high frequencies. The first two characteristics appear in the *dizi* tones in the second and third registers, while the last one appears in the *dizi* tones in the third register.

In section 5.3, the deriving of the relation between the lowest three harmonics generated by a non-linear oscillator is restricted to the frequency-domain. This method of harmonic balance is

mainly applied to small- and medium-amplitude oscillations. To explore the membrane's behavior in large-amplitude oscillations, the time-domain approach is more suited than the frequency-domain approach.

In the time-domain, numerical simulations of the membrane's motion are performed by modeling it as a Duffing oscillator driven by a quasi-sinusoidal force. The membrane's equation of motion is

$$\ddot{x} + r\dot{x} + w_m^2 x + bx^3 = \frac{G_0}{m}\cos(wt + j_0) + \frac{G_1}{m}\cos(2wt + j_1) + \frac{G_2}{m}\cos(wt/2 + j_2) \qquad (6.1)$$

with

$$b > 0, \quad G_0 >> G_1 \text{ and } G_0 >> G_2 .$$

As the force terms corresponding to the second harmonic and the first subharmonic are assumed to be much smaller than that corresponding to the fundamental, the force driving the membrane is quasi-sinusoidal.

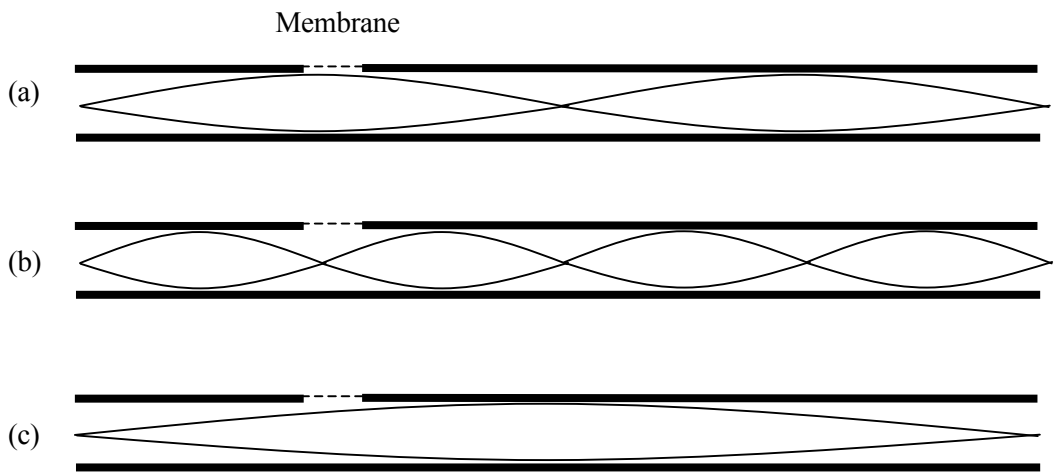


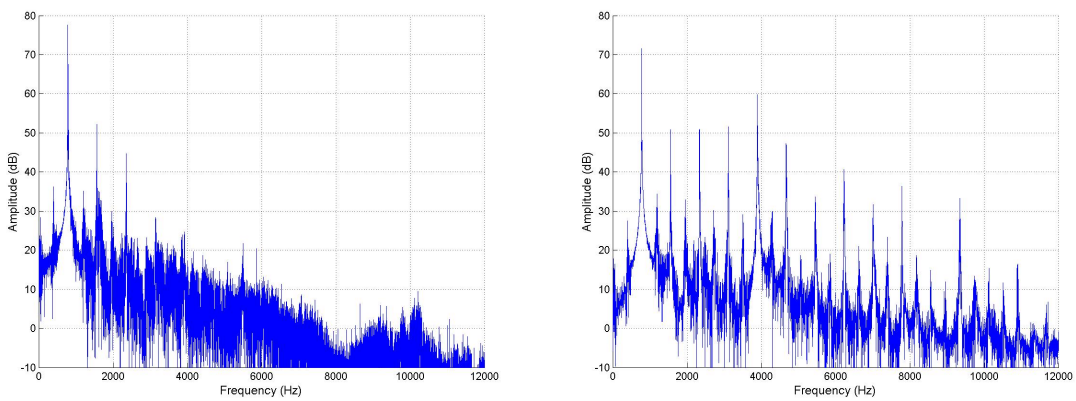
Figure 6.4: *Acoustic pressure waves in the dizi for a tone in the third register. (a) Fundamental wave corresponding to the excitement of the second mode. (b) Second harmonic wave. The membrane is located near a pressure node of its acoustic wave. (c) Subharmonic wave corresponding to the excitement of the first mode.*

Eq. (6.1) is available for simulating the *dizi* tones in the second and third registers. For the *dizi* tones in the third register, it is assumed that $G_1/G_0 = o(10^{-1})$ and $G_2/G_0 = o(10^{-1})$. The relatively small amplitude of the force term corresponding to the second harmonic is due to the membrane's location. In the second and third registers, the membrane is located near a pressure antinode of the fundamental acoustic wave [Fig. 6.4(a)], and a node of the second harmonic wave [Fig. 6.4(b)].

Therefore, if these *dizi* tones are produced by a normal embouchure, for which the jet vibration is approximately symmetric, the acoustic pressure under the membrane is dominated by the fundamental. On the other hand, for the *dizi* tones in the first register the acoustic pressure of the second harmonic has larger amplitudes at the membrane's position; $G_1 \sim G_2$. So the driving force is no longer quasi-sinusoidal and the spectrum shows no predominance of odd-numbered harmonics. The predominance of odd-numbered harmonics in the *dizi* tones in the second and third registers supports that the non-linear behavior of the membrane is predominated by the cubic non-linearity.

The subharmonic force term in Eq. (6.1) stem from the possible excitement of the first mode of the air column when the *dizi* is played in the second mode. The coexistence of the first and second modes results in multi-phonic flute tones, and weak subharmonic components with frequencies $(2n\text{-}1)f_0/2$ can be observed in the spectrum. In the simulations, only the contribution of the first subharmonic is considered, as it is presumed that the acoustic field in the pipe is predominated by low frequencies.

The subharmonic generation of a membrane-less flute differs from that of a membrane flute. Two flute tones with subharmonics are compared in Fig. 6.5. A membrane-less flute can generate the first, second and third subharmonics, while a membrane flute can generate upper subharmonics. At high frequencies, these subhamonics are comparable to the flanking harmonics in magnitude.



(a) (b)

Figure 6.5: *Comparison of the subharmonic generation of a membrane-less and a membrane flute in the second mode. Sounds were produced by the large dizi in G with the fingering* ●●●○○○○. *(a) Dizi tone spectrum for a membrane covered by the palm of a hand. (b) Dizi tone spectrum for a vibrating membrane.*

In this section, the presence of upper subharmonics of a *dizi* tones in the third register (second mode) is expected to explain by numerical simulations with $G_2 > 0$ in Eq. (6.1). For simulating the *dizi* tones in the second register (in the first play mode), it is assumed that $G_2 = 0$ because no lower mode of the pipe can be excited.

Although both the quadric and cubic non-linearities of the membrane were demonstrated in

chapter 5, only the cubic term is considered in Eq. (6.1). From a theoretical view, the ignorance of quadric non-linearity is due to the large amplitude of the membrane's vibration [see Eq. (5.8)].

The ignorance of the quadric non-linearity is supported by the predominance of odd-numbered harmonics for the *dizi* tones in the second and third registers, for which the membrane is located near a pressure node of the second harmonic. When the membrane is driven by a sinusoidal force, the cubic non-linearity only generates odd-numbered harmonics, while the quadric non-linearity generates both odd- and even-numbered harmonics. The predominance of odd-numbered harmonics in the *dizi* tones in the second and third registers leads to a quasi-sinusoidally driven Duffing oscillator of the membrane.

6.3 Methods

6.3.1 Experimental study

6.3.1.1 Part one: Harmonic generation

The first part of the experiment is dedicated to the membrane's harmonic generation for the *dizi* tone in the second register. Fig. 6.6(a) shows the general features of the experimental arrangement. In order to measure the sounds radiated by the membrane, a microphone was placed approximately 1 cm above the membrane. A box enclosed the membrane so as to insulate the microphone from sounds radiated at the embouchure hole, finger holes and from the environment. The instrument used was a large *dizi* in G and the fingering is ●●○○○○○. Its physical dimensions are presented in Fig. 6.6(b).

Two kinds of tones driving the membrane were produced and recorded: *dizi* tones produced by blowing the instrument and tones produced by singing into the *dizi*. Singing into the *dizi* is an innovative performance technique of the *dizi*. In this experiment, this technique was used with falsetto voices to provide an alternative quasi-sinusoidal sound source with a similar fundamental frequency of the *dizi* tones in the second register.

There are two reasons to use singing voices for driving the membrane. First, singing into the *dizi* can produce acoustic waves in the pipe whose intensity is smaller than *pianissimo dizi* tones. In chapters 4 and 5, the intensity of signals applied to the *dizi* is much weaker than the acoustic waves in the *dizi* pipe in normal playing states. For exploring the non-linear behavior of the membrane driven by signals with various intensities, singing into the *dizi* can provide a transition from weak signals generated by speakers to strong signals generated by blowing the instrument. Second, producing tones by singing into the *dizi* can avoid the complicated aerodynamics at the mouth involved in *dizi* sounding. Harmonics produced by the membrane may affect the jet motion at the mouth. On the other hand, for tones produced by singing into the *dizi*, it is presumed that the effect of the high harmonics to the glottis is negligible. Falsetto was used for singing into the *dizi* because it produces

weaker harmonics than the modal register. Hence, the glottis can be approximated as a source of quasi-sinusoidal tones. When tones were produced by singing into the *dizi*, the mouth hole was completely covered by the lips. So the *dizi* combined with the vocal tract and an open-close tube was formed, as illustrated in Fig. 6.6(c) with a simplified geometry.

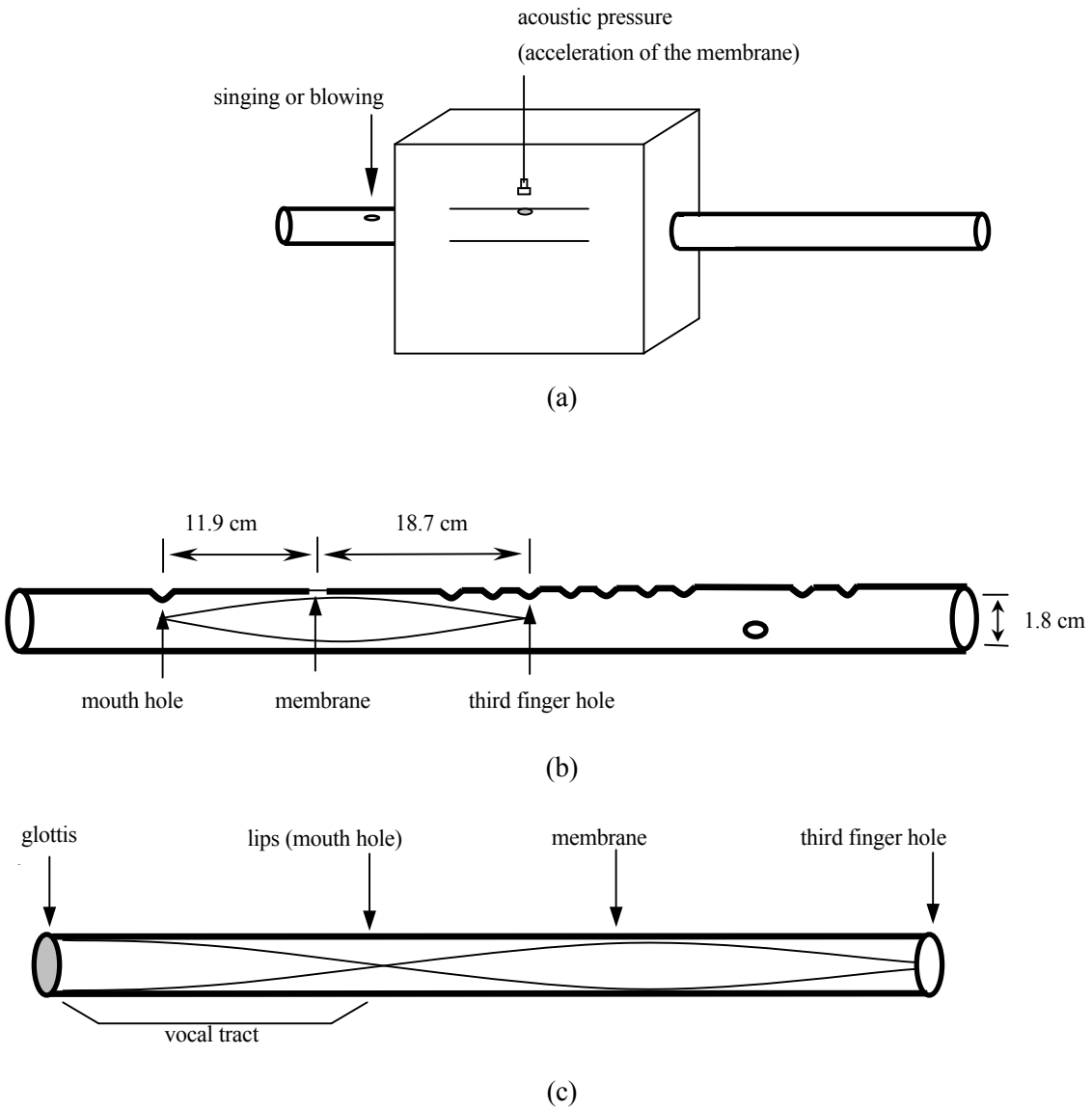


Figure 6.6: *Experimental setup for membrane's harmonic generation. (a) Experimental setup. (b) Fundamental acoustic pressure waves in the large dizi in G for the tones in the second registers. Fingering ●●○○○○○. (c) Fundamental acoustic pressure waves in the large dizi in G induced by singing into the dizi. Fingering ●●○○○○○. A vocal-tract-dizi tube is formed by covering the mouth hole with lips. In this figure the irregular shape of the vocal tract is approximated as a cylinder with the same diameter of the dizi.*

As the tones produced by singing into the *dizi* are limited in intensity, real *dizi* tones produced by blowing must be used to induce larger-amplitude vibrations of the membrane. It should be noted that although the *dizi* tones could reveal the membrane's behavior under strong forcing, the non-linear aerodynamics at the mouth could interact with the membrane through upper harmonics propagating in the pipe. This makes the right-hand side of Eq. (6.1) unrealistic and introduces complications that are beyond the considerations of the present simple model.

The acoustic pressure waves recorded 1 cm above the membrane is approximately in proportion to the acceleration of the membrane (see section 5.4.3). They were integrated twice to give the membrane's displacement. The initial velocity was chosen to give a quasi-periodic displacement of the membrane.

6.3.1.2 Part two: Subharmonic generation

The second part of the experiment was dedicated to the membrane's subharmonic generation for the *dizi* tones in the third register (second play mode). The instruments used were a large *dizi* in G with the fingering ●●●●●●○, and a *qudi* in D with the fingering ●●●○○○. Their physical dimensions are presented in Fig. 6.7.

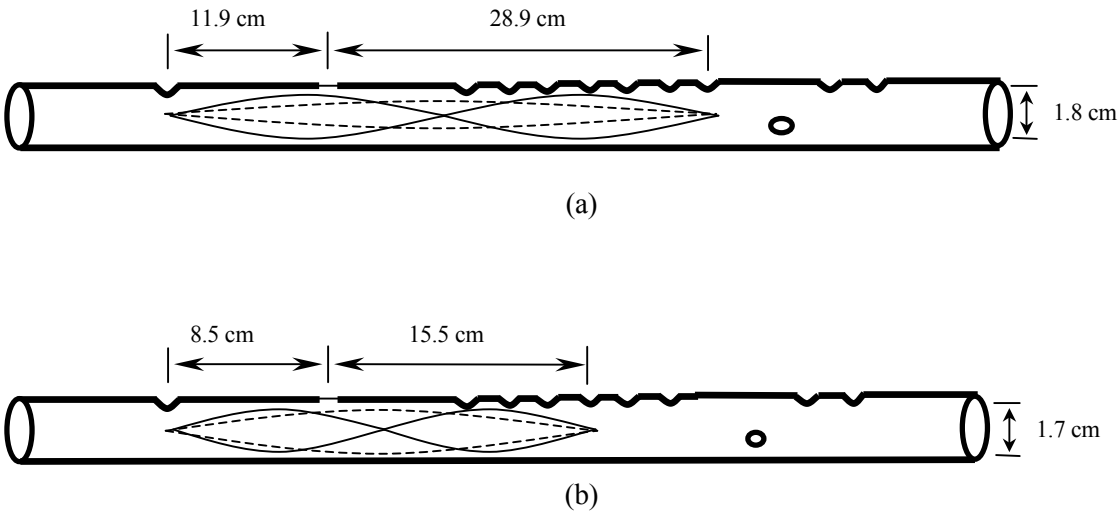


Figure 6.7*: Experimental setup for membrane's subharmonic generation. Comparison between the fundamental (—) and the first subharmonic (– –) acoustic pressure waves. (a) Dizi tone in the third register of the large dizi in G. Fingering* ●●●●●●○*. (a) Dizi tone in the third register of the qudi in D. Fingering* ●●●○○○*.*

Although singing into the *dizi* can avoid the involvement of the aerodynamics at the mouth hole, it cannot be used for exploring the membrane's subharmonic generation. Subharmonics generated by

singing are difficult to control and not restricted to low frequencies. This makes the right-hand side of Eq. (6.1) unrealistic in describing the signals produced by voices with subharmonics. Consequently, the membrane's subharmonic generation was studied with the real *dizi* tones in the third register.

Two instruments and two fingerings were used for producing different subharmonics: the large *dizi* in G with the fingering ●●●●●●○ and the *qudi* in D with the fingering ●●●○○○. The latter can produce subharmonics whose spectral lines are widened (see sections 3.4 and 4.4). It is suggested in section 4.4 that this spectral lines widening may stem from the stretched octave; the resonance frequency of the first mode of the pipe deviates from $f_0/2$ (f_0 is the resonance frequency of the second mode).

6.3.2 Numerical study

In numerical simulations, the terms in Eq. (6.1) is discretized in time with a forward Euler method using time step Δt. This results in

$$\begin{aligned}&\frac{v_{n+1}-v_n}{\Delta t}+rv_n+\omega_m^{\,2}x_n+bx_n^{\,3}\\&=\frac{G_0}{m}\cos(\omega t_n+\phi_0)+\frac{G_1}{m}\cos(2\omega t_n+\phi_1)+\frac{G_2}{m}\cos(\omega t_n/2+\phi_2)\end{aligned} \tag{6.2}$$

with

$$v_n=\frac{x_{n+1}-x_n}{\Delta t} \tag{6.3}$$

and

$$\Delta t=10^{-5}\ \text{sec}\,.$$

This value of Δt had been demonstrated to be small enough to give smooth and periodic phase plots of the membrane. The values of the effective acoustic mass m, the damping coefficient $r = R/m$ and the angular resonant frequency ω_m of the membrane were chosen to fit the response curve of the membrane measured in the impedance experiments (chapter 4). The cubic non-linearity, the magnitudes and the phases in the forcing terms were determined by to fit the experimental results. The initial velocity is chosen to give a quasi-periodic displacement ξ_n. It should be noted that although the numerical approach aims to explain the spectral features of *dizi* tones, the experimental results and the numerical results are primarily compared in the time-domain. The waveforms of sounds radiated by the membrane and the membrane's phase plots are expected to reveal the performance of the model.

6.4 Comparisons between numerical and experimental results

6.4.1 Harmonic generation

All simulations of Eq. (6.1) were performed with the membrane's damping coefficient r = 1700 sec^{-1}, resonant frequency $\omega_m/2\pi = f_m$ = 2500 Hz, effective acoustic mass $m = 5\times10^{-7}$ kg. By fitting the experimental results, the coefficient of cubic non-linearity was estimated as $\beta = 10^{17} m^{-2} sec^{-2}$.

Figs. 6.8 to 6.12 show the five acoustic waves in the *dizi*. Three of them were produced by singing into the *dizi*, while the other two large-amplitude waves were produced by playing the instrument. As the microphone cannot measure absolute values of the acoustic pressure, the experimental results of the membrane's displacement, velocity and acceleration are plotted in arbitrary unit. In the numerical results, they are plotted in MKS system. The displacement of the membrane is of the order of 10^{-5} m. Fitting the experimental results yields values of G_0 in Eq. (6.1). The acoustic pressure driving the membrane G_0/S_m increased from 50 Pa to 150 Pa (Figs. 6.8–6.12). These values of acoustic pressure in the *dizi* are in agreement with previous measurements of the acoustic pressure in flue instruments (e.g., Ségoufin *et al.* 2000, Verge *et al.* 1997, Verge *et al.* 1994).

The experimental results of tones produced by singing into the *dizi* are successfully simulated by a quasi-sinusoidally forced Duffing oscillator (Figs. 6.8–6.10). The membrane's phase plots show increasing interlocking spirals, corresponding to the increasing fluctuations at the two extremes of the membrane's displacement waveforms. The fact that the center frequency of the first formant of the *dizi* tones increases with the tone intensity is also successfully predicted by the simulations of this model.

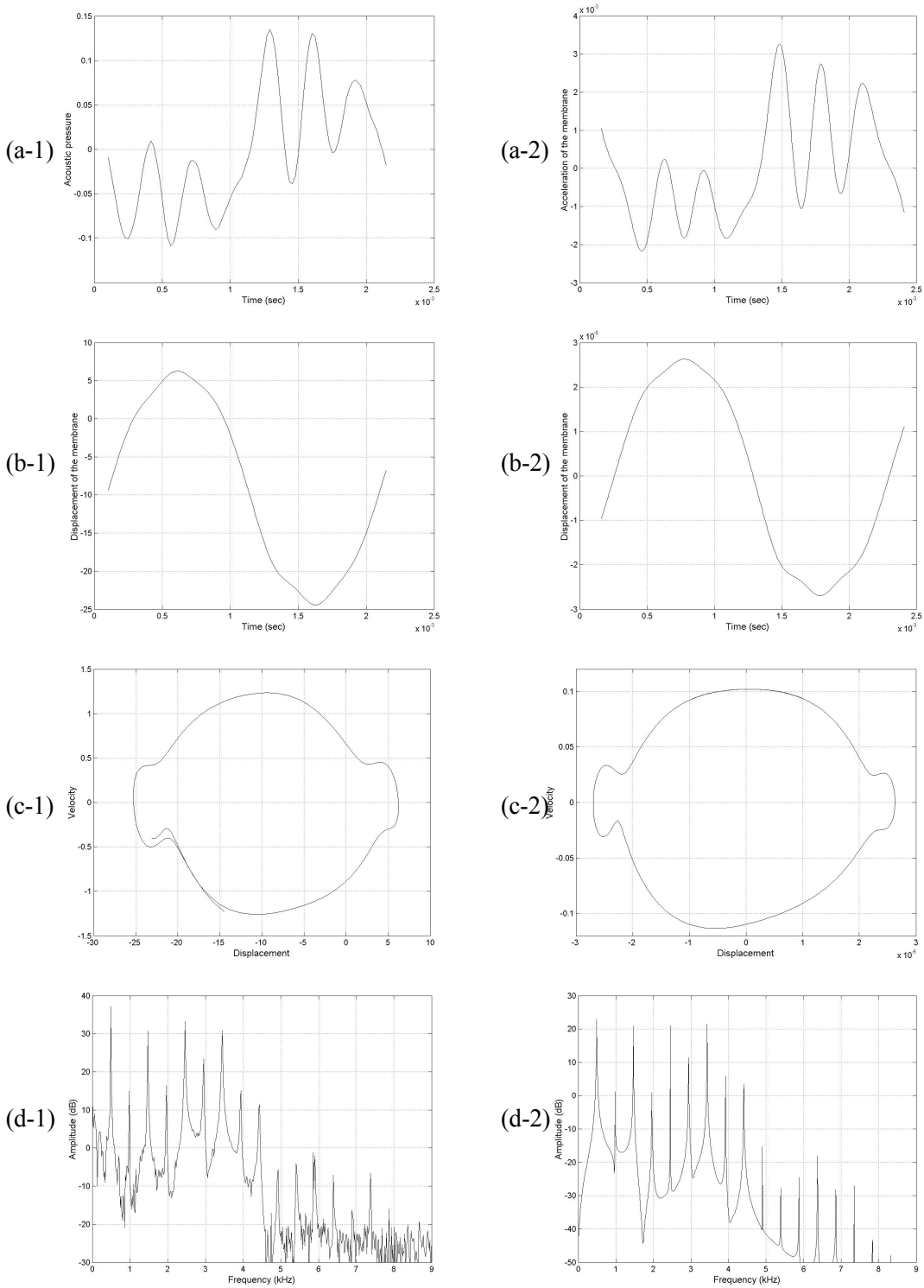


Figure 6.8: *Comparison between the tone radiated by the membrane produced by singing into the dizi (left) and numerical results (right). Stimulation of Eq. (6.1) with* $G_0 = 4\times10^{-3} m \cdot sec^{-2}$, $G_1 = .02G_0$, $G_2 = 0$, $\omega/2\pi = 490Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1 = 1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

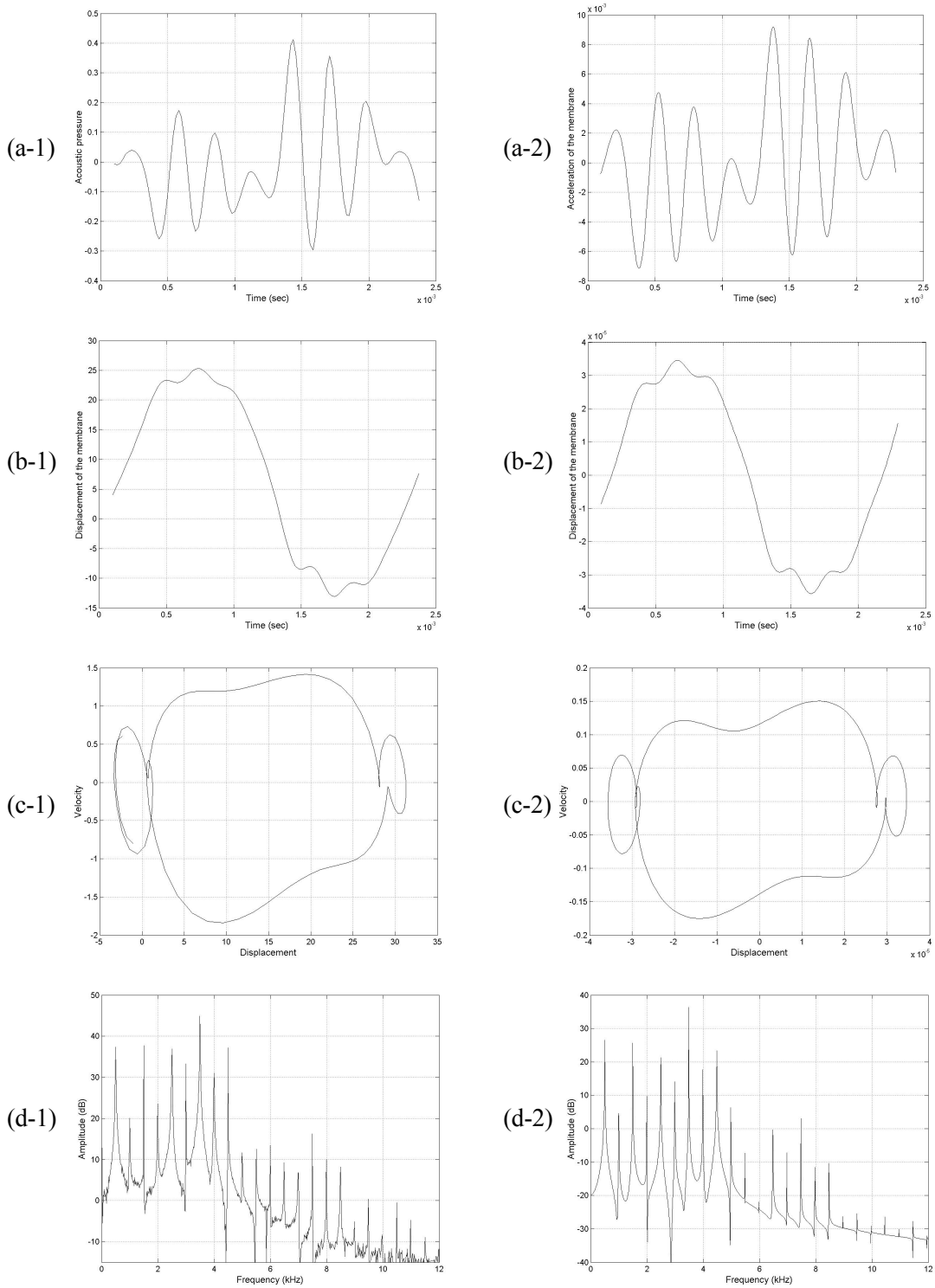


Figure 6.9: *Comparison between the tone radiated by the membrane produced by singing into the dizi (left) and numerical results (right). Stimulation of Eq. (6.1) with* $G_0 = 5.75\times10^{-3} m \cdot sec^{-2}$, $G_1 = .03G_0$, $G_2 = 0$, $\omega/2\pi = 498Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1 = 1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

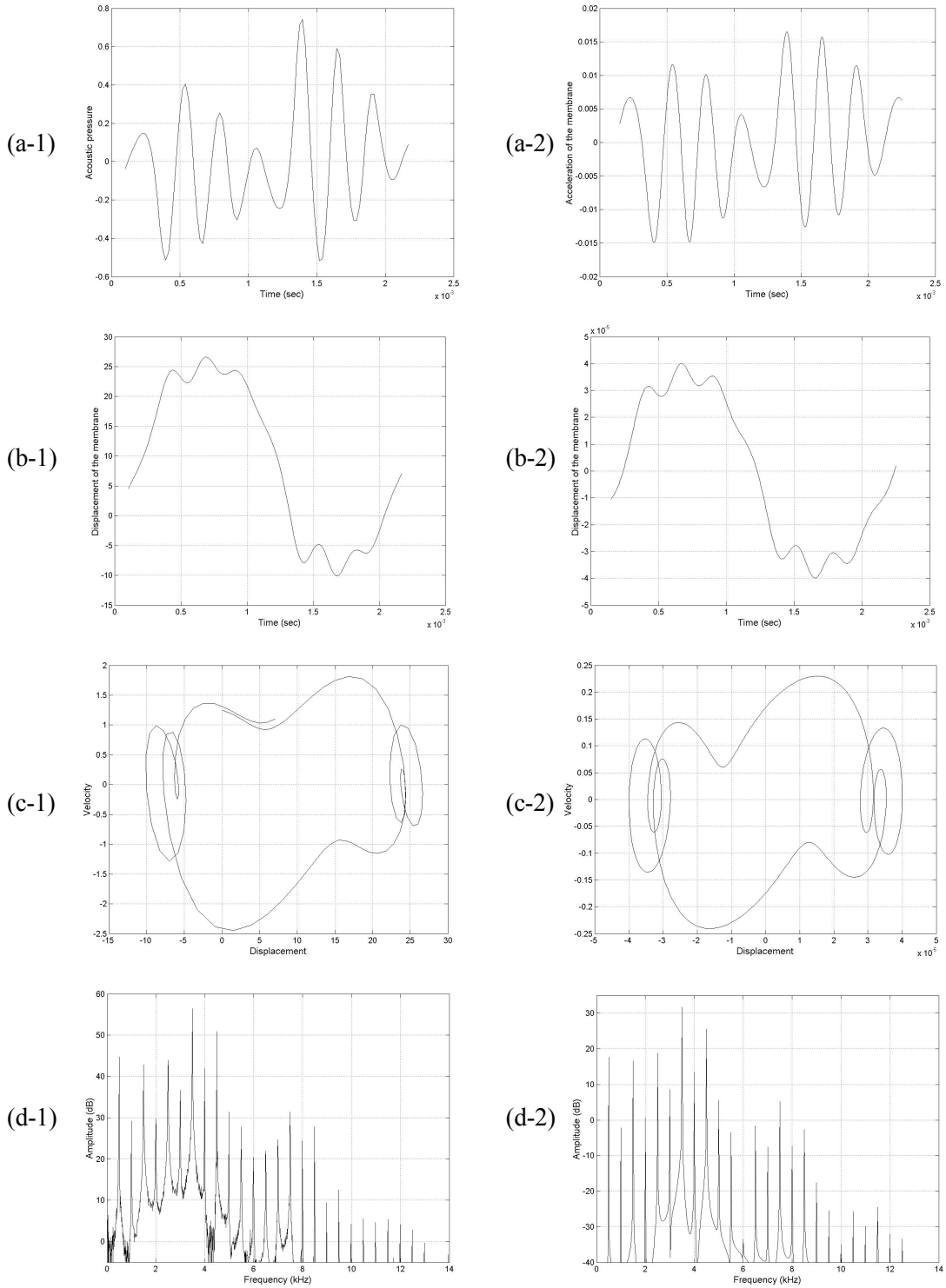


Figure 6.10: *Comparison between the tone radiated by the membrane produced by singing into the dizi (left) and numerical results (right). Stimulation of Eq. (6.1) with* $G_0 = 6.65\times10^{-3} m \cdot sec^{-2}$, $G_1 = .03G_0$, $G_2 = 0$, $\omega/2\pi = 500Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1 = 1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

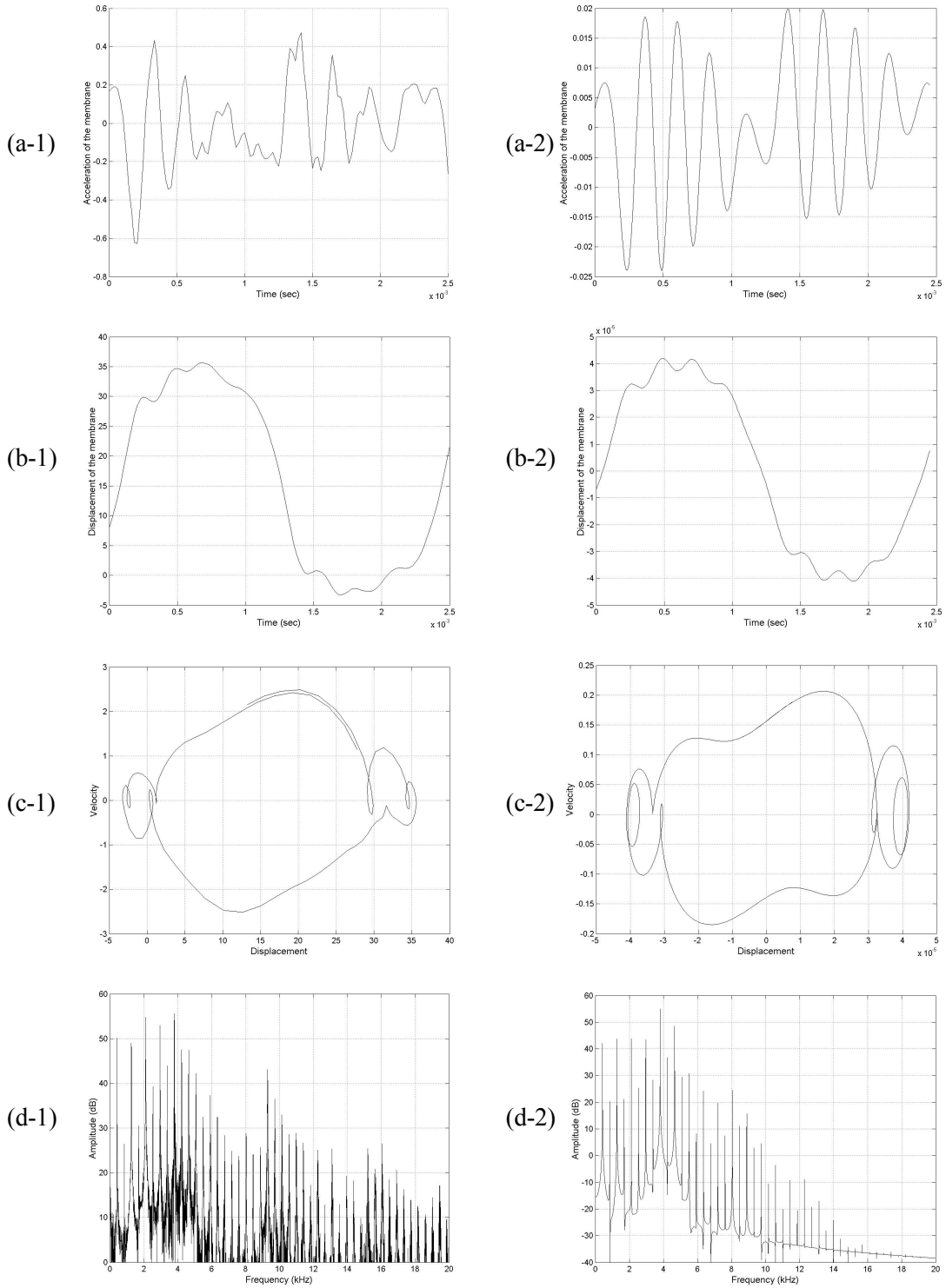


Figure 6.11: *Comparison between the tone radiated by the membrane produced by playing the dizi (left) and numerical results (right). Stimulation of Eq. (6.1) with* $G_0 = 8\times10^{-3} m \cdot sec^{-2}$, $G_1 = .02G_0$, $G_2 = 0$, $\omega/2\pi = 424Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1=1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

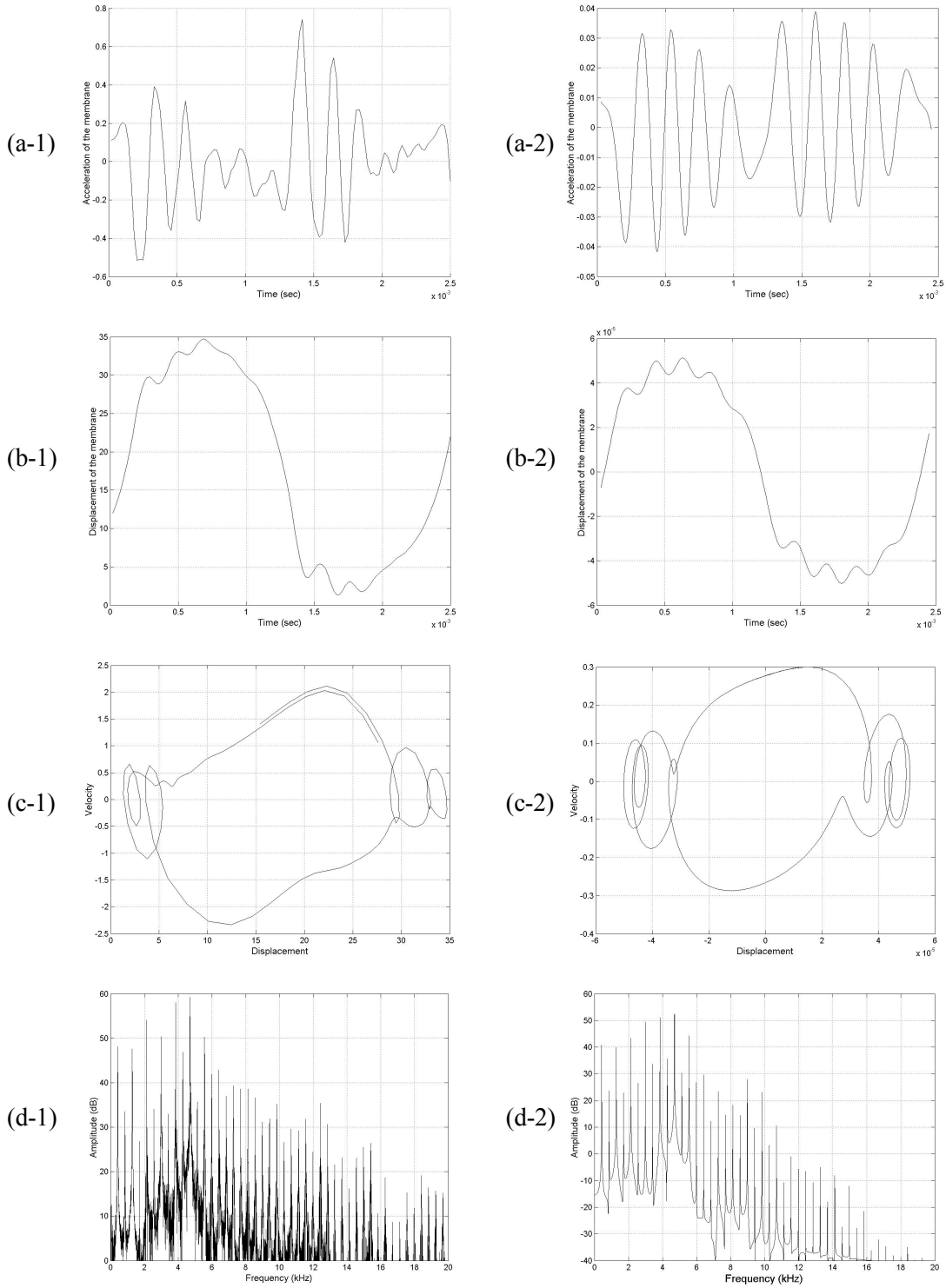


Figure 6.12: *Comparison between the tone radiated by the membrane produced by playing the dizi (left) and numerical results (right). Stimulation of Eq. (6.1) with* $G_0 = 1.1\times10^{-2} m\cdot sec^{-2}$, $G_1 = .05G_0$, $G_2 = 0$, $\omega/2\pi = 428.1Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1 = 1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

For the tones produced by playing the *dizi* the agreement is restricted to frequencies lower than 10 kHz (Figs. 6.11 and 6.12). Although the membrane's phase plots of experimental and numerical results show similar interlocking spirals, the simulated tones do not contain rich harmonics above 10 kHz. This discrepancy arises the possibility that either the left-hand side or the right-hand side of Eq. (6.1) is unrealistic.

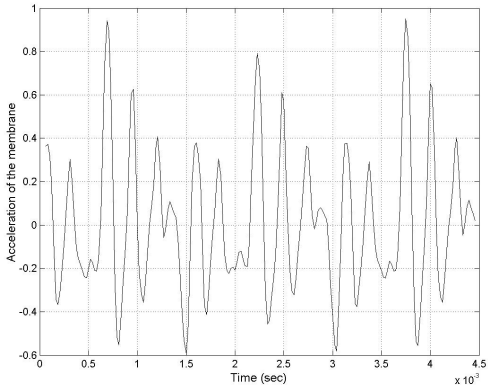
The left-hand side of Eq. (6.1) take into account neither the non-linear terms of orders higher than cubic nor the non-linear damping. When higher non-linear terms are added to it, however, I found that the membrane still cannot generate rich harmonics above 10 kHz. Note that the tone produced by singing into the *dizi* shown in Fig. 6.10 has a spectral content very similar to the *dizi* tone shown in Fig. 6.11 at frequencies below 5 kHz. So the vibration amplitudes of the membrane were approximately the same for these two tones. As the higher-order non-linearity and the nonlinear damping are not considered for the tone produced by singing into the *dizi* shown in Fig. 6.10, they may also not contribute to the *dizi* tone shown in Fig. 6.11.

The driving force in the right-hand side of Eq. (6.1) contains the fundamental and the second harmonic acoustic waves in the pipe. This approximation is not suitable for loud *dizi* tones, where higher harmonics generated by the membrane are non-ignorable. However, rich harmonics above 5 kHz are not observed in the tone spectrum produced by singing into the *dizi* [Fig. 6.10(d-1)], but are noticeable in the *dizi* tone spectrum [Fig. 6.11(d-1)]. Therefore, the rich harmonics above 5 kHz in the pipe are likely to stem from the non-linear aerodynamics at the mouth hole. Note that when the membrane is located near a pressure node of the fundamental wave, the *dizi* tones contain no components above 10 kHz. Therefore, neither the membrane nor the aerodynamics at the mouth hole alone is able to produce harmonics above 10 kHz. But when both non-linear mechanisms exist, their interaction could generate harmonics above 10 kHz. If the jet oscillation is sensitive to the harmonics around 5 kHz generated by the membrane, it may induce equivalent pressure fluctuations with rich harmonics, and the amplified harmonics around 5 kHz may drive the membrane to radiate harmonics above 10 kHz. The issue of the interactions between the membrane and the jet through upper harmonics propagating in the pipe will be discussed in section 6.6.

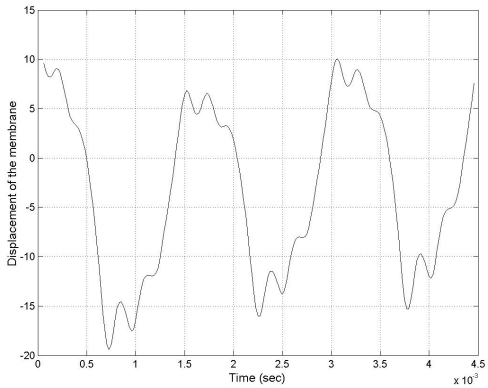
In the study of the membrane's harmonic generation, the experimental results suggest that the first formant of *dizi* tones can be explained by modeling the membrane as a quasi-sinusoidally driven Duffing oscillator, but the harmonics above 10 kHz cannot be predicted without considering higher harmonic terms of the driving force in Eq. (6.1).

6.4.2 Subharmonic generation

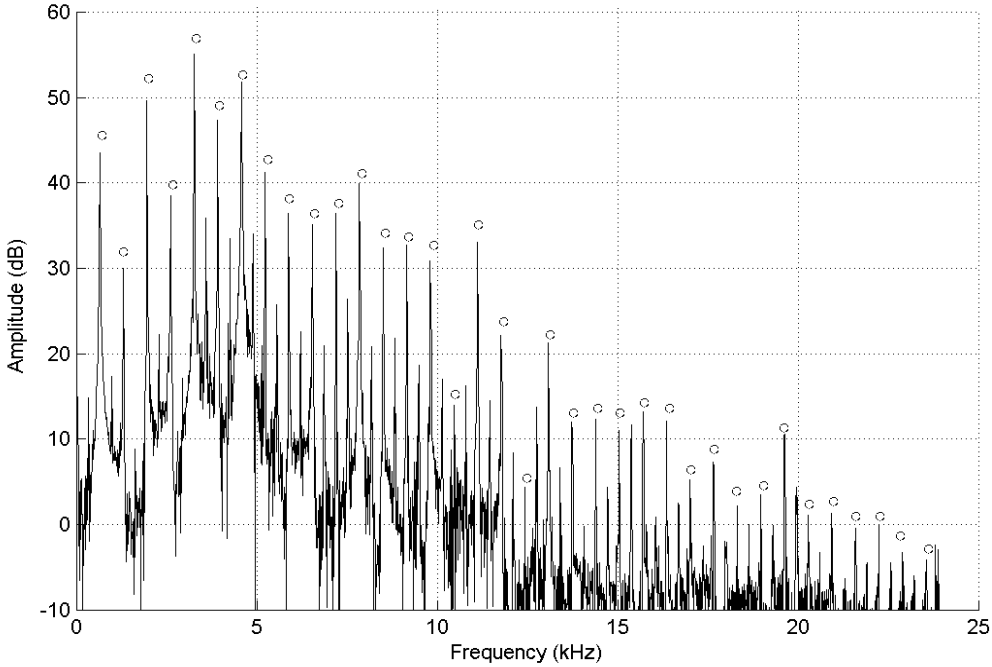
Fig. 6.13 presents the waveforms of the membrane and spectrum of the *dizi* tone containing subharmonics. A doubled period in the waveform of the acoustic pressure can be observed in Fig. 6.13(a). By integrating twice, however, the displacement waveform [Fig. 6.13(b)] loses this periodicity because the sound is quasi-periodic. Due to fluctuations the membrane's phase plots cannot be reconstructed.



(a)



(b)



(c)

Figure 6.13: *Dizi tone with subharmonics with frequencies $(2n-1)f_0/2$. The tone was recorded above the dizi membrane when a large dizi in G was played in the second play mode with the fingering* ●●●●●●○. *(a) Tone radiated by the membrane. (b) Displacement. (c) Spectrum of (a). The spectral lines of harmonics are marked with circles.*

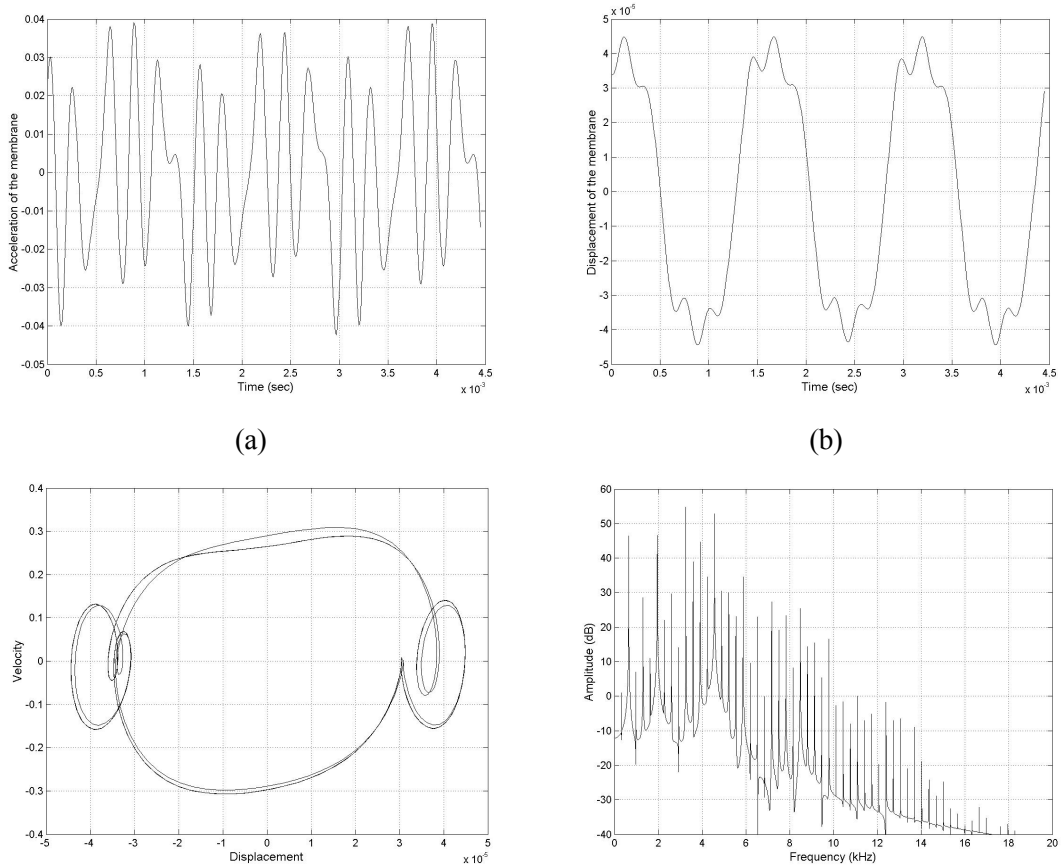


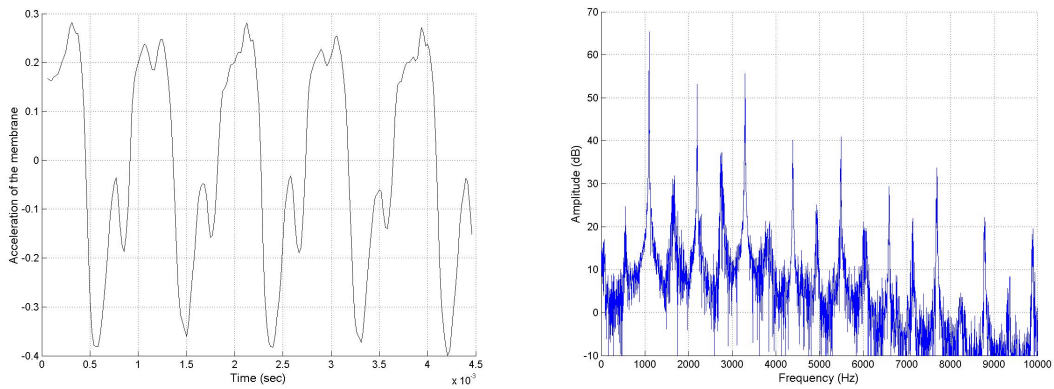
Figure 6.14: *Stimulation of Eq. (6.1) with* $G_0 = 8\times10^{-3} m \cdot sec^{-2}$, $G_1 = .05G_0$, $G_2 = .02G_0$, $\omega/2\pi = 653Hz$, $\varphi_0 = \varphi_2 = 0$, $\varphi_1 = 1$. *(a) Tone radiated from the membrane. (b) Displacement. (c) Phase plots. (d) Spectra of (a).*

Fig. 6.14 shows the simulated results of a *dizi* tone containing subharmonics. A period-doubling can be seen in the simulated phase plots [Fig. 6.14(c)]. Although the membrane's displacement waveforms obtained in the experiment [Fig. 6.13(b)] and the simulation [Fig. 6.14(b)] are qualitatively similar, a comparison of the membrane's acceleration waveforms [Fig. 6.13(a), Fig. 6.14(a)] and their spectra [Fig. 6.13(c), Fig. 6.14(d)] shows that the simulated *dizi* tone lacks rich components above 10 kHz. This discrepancy is attributed to the jet's sensitivity to high frequencies (see section 6.6).

From a perceptual view, the most important feature of the subharmonic generation of the *dizi* is that the lower subharmonics are much weaker than the adjacent harmonics, while the upper subharmonics are comparable to the adjacent harmonics in amplitude. This feature can be observed in Fig. 6.13(c), where the subharmonics below 3 kHz are approximately 30 dB smaller than the adjacent harmonics, and the subharmonics around 8 kHz are approximately 10 dB smaller than the adjacent harmonics. The simulated spectrum [Fig. 6.14(d)] shows amplitude differences between

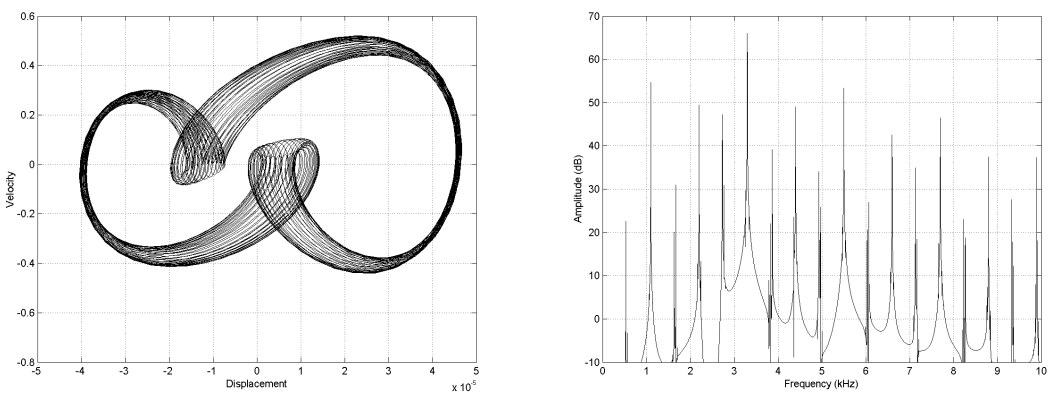
harmonics and subharmonics at low frequencies similar to those in the recorded spectrum [Fig. 6.13(c)].

Another *dizi* tone was recorded for exploring the phenomenon of spectral lines widening of subharmonics, which can be observed in the highest two notes of the third register of the *qudi* and the *bangdi*. Fig. 6.15 shows the waveform and the spectrum of such a tone produced by a *qudi* in D. This phenomenon is simulated by setting the subharmonic force term with a frequency deviating from $f_0/2$. When the frequencies of the harmonic forcing and the subharmonic forcing stand in the integer ratio 2:1, the membrane oscillation is periodic. But for the fingerings ●●●●○○ and ●●●○○○, these two frequencies do not exactly stand in the integer ratio 2:1 (see section 4.4.2). The multi-phonic produced by simultaneously exciting the first and second resonances is not exactly an octave.



(a) (b)

Figure 6.15: *Dizi tone with subharmonics whose spectral lines are widened. The tone was recorded above the membrane when a qudi in D was played in the second mode with the fingering* ●●●●○○. *(a) Tone radiated from the membrane. (b) Spectrum of (a).*



(a) (b)

Figure 6.16: *Simulation of spectral-line-widening of subharmonics. The angular frequency of the subharmonic forcingin Eq. (6.1) is* $\omega/2.07$, *with* $\omega/2\pi = 1100$*Hz.* $\varphi_0 = 0$, $\varphi_1 = 1$, $\varphi_2 = -1$, $G_0 = 4.5 \times 10^{-4} m \cdot sec^{-2}$, $G_1 = .1G_0$, $G_2 = .05G_0$. *(a) Phase plots. (b) Spectrum of the tone.*

Fig. 6.16 presents the phase plots and spectrum of a *dizi* tone with widened spectral lines of subharmonics. The frequencies of harmonic forcing and subharmonic forcing are set to be in the ratio 2.07:1, which mimics the stretched octave of 1257 cents for B4/B5 (the *qudi* in C with the fingering ●●●●○○) in Fig. 4.17. The simulated phase plots of the membrane show a quasi-periodicity [Fig. 6.16(a)] and the simulated spectrum shows split spectral lines of subharmonics [Fig. 6.16(b)]. This partially explains the widened spectral lines of subharmonic.

6.5 Membrane location and the *dizi* timbre

In principle, the Duffing oscillator model of the membrane is able to explain the three major spectral features of the *dizi* tones in the second and third registers: the first formant, the odd-numbered harmonics predominance, and the subharmonics at high frequencies. Although this model fails to predict the spectral components above 10 kHz of *dizi* tones, its performance is good for soft *dizi* tones. For loud *dizi* tones, the acoustic field in the pipe is no longer dominated by low frequencies and the membrane is not driven by a quasi-sinusoidal force.

Despite the crudeness of the Duffing oscillator model of the membrane, the present results could be generalized to relate the inhomogeneity of the *dizi* timbre in the five registers to the location of the membrane. Different registers correspond to different location relations between the membrane and the standing pressure waves in the *dizi*. Fig. 6.17 shows the membrane's relative location to the standing waves of the fundamental, the second harmonic and the first subharmonic in the pipe for five registers. The spectral features of the five registers of the *dizi* are summarized in Tab. 6.1.

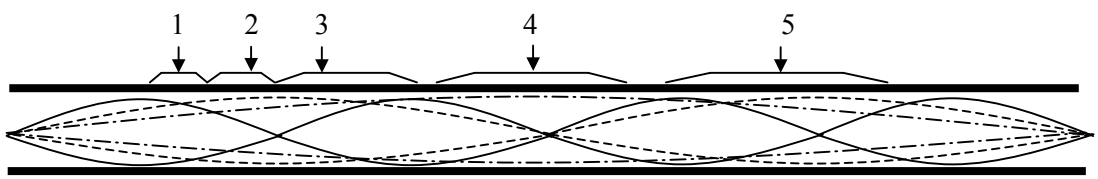


Figure 6.17: *Membrane's relative location to the standing waves of the fundamental (– –) , the second harmonic (—) and the first subharmonic (· —) in the pipe. For a real dizi, their wavelengths vary with the tone pitch. This is analogized here by taking their wavelengths constant and the membrane's location variable. The number indicates the index of the register and the brace indicates the range of the membrane's location.*

<table>
<tr><th rowspan="2">Register</th><th rowspan="2">Timbre characteristics</th><th colspan="3">Location of the dizi membrane</th></tr>
<tr><th>Fundamental</th><th>Second harmonic</th><th>First subharmonic</th></tr>
<tr><td>1</td><td>Rich harmonics</td><td>Between node and antinode</td><td>Near antinode</td><td rowspan="2">Amplitude = 0</td></tr>
<tr><td>2</td><td>Rich harmonics, odd-numbered harmonics predominance</td><td rowspan="2">Near antinode</td><td rowspan="2">Near node</td></tr>
<tr><td>3</td><td>Rich harmonics, Subharmonics (odd-numbered harmonics predominance)</td><td rowspan="2">Depending on notes</td><td>Between node and antinode</td></tr>
<tr><td>4</td><td>-</td><td>Near node</td><td>Near antinode</td></tr>
<tr><td>5</td><td>Rich harmonics</td><td>Near antinode</td><td>Between node and antinode</td></tr>
</table>

Table 6.2: *Membrane's location correlates of spectral features of the five registers of the dizi. Shadowed areas represent the physical mechanisms corresponding to the spectral features.*

In the first register, the membrane is located near a pressure antinode of the acoustic wave of the second harmonic, thus driven by both the fundamental and the second harmonic; the radiating sounds contain both odd-numbered and even-numbered harmonics. In the second register, the membrane is located near a pressure node of the second harmonic wave, thus driven by the dominant fundamental; the radiating sounds are predominated by odd-numbered harmonics. This spectral feature also holds for lower notes in the third register and becomes less noticeable for upper notes in this register because the membrane is no longer located near a pressure node of the second harmonic wave.

The present model of the *dizi* membrane also predicts that if the jet oscillations are so asymmetric that the second harmonic is stronger than the fundamental, the spectrum of the *dizi* tone in the first register will show a predominance of the even-numbered harmonics, because the acoustic pressure driving the membrane is predominated by the second harmonic. This spectral feature has been observed in the lowest *dizi* tone (Fig. 6.18).

Subharmonics at high frequencies characteristic of the *dizi* tones in the third register stem from the subharmonic force $G_2\cos(\omega t/2)$ exerting on the membrane, which is due to the excitement of the first mode of the air column. For a membrane-less flute, the subharmonic waves $(2n\text{-}1)f_0/2$ are restricted to low frequencies and tend to be masked by adjacent strong harmonics. For a membrane flute, high-frequency subharmonics are generated by the non-linear membrane under the weak subharmonic force $G_2\cos(\omega t/2)$ combined with the strong fundamental force $G_0\cos(\omega t)$.

Consequently, subharmonics at high frequencies are comparable in amplitude with adjacent harmonics.

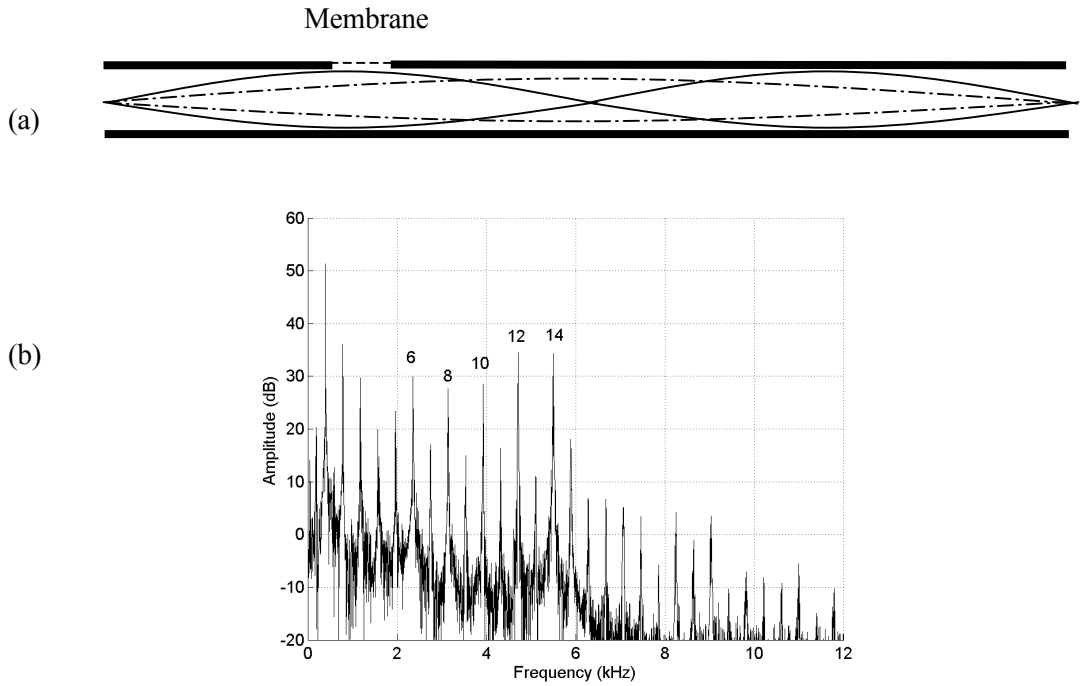


Figure 6.18: *Predominance of the even-numbered harmonics in the lowest dizi tone. (a) Sketch of the pressure waves of the fundamental (· —) and the second harmonic (—). Due to asymmetric oscillations of the jet, the second harmonic is stronger than the fundamental. (b) Spectrum with dominant even-numbered harmonics around the first formant.*

The dull *dizi* tones in the fourth register and the bright *dizi* tones in the fifth register can easily be explained by the membrane's location (see section 3.3). It should be noted that although the three spectral features of the *dizi* tones are successfully explained by a Duffing oscillator model of the *dizi* membrane, the present experiments merely account for the *dizi* tones in the second and third registers. Further studies should involve *dizi* tones in all registers and measurements of the absolute values of the acoustic pressure in the *dizi*.

6.6 Aerodynamics and radiation at the mouth hole

6.6.1 Drive mechanism of flue instruments

In this last section on the physics of the *dizi*, the difficult problem of the aerodynamics at the mouth hole is theoretically discussed. As there have been a lot of investigations on the physics of flue instruments (for a recent review, see Fabre *et al.* 2000), this section mainly concerns the possible interactions between the non-linear aerodynamics and the non-linear membrane of a membrane-flue

instrument.

A series of progress of physical modeling flue instruments has been made in the past decade by physicists at the Eindhoven University of Technology in the Netherlands (with cooperation of Institut de Recherche et de Coordination Acoustique et Musique Paris). Their models were based on sophisticated theories of fluid mechanics such as Lighthill's definition of the acoustic field and source (Lighthill 1954), potential theory (Howe 1984), and vortex-sound theory (Powell 1964), supported by flow visualizations and dimensionless representations (Verge *et al.* 1994a, Verge *et al.* 1994b, Verge *et al.* 1997a, Verge *et al.* 1997b, Fabre *et al.* 1996). Dequand's (2000) model distinguished two sound source models according to the mouth geometry of flue instruments: the jet-drive model and the discrete-vortex model. The jet-drive model qualitatively explained the drive mechanism of most flutes, recorders and so on, where the ratio of the mouth width W and the jet thickness h is large; $W/h > 2$ (see Fig. 6.19). On the other hand, for $W/h < 2$ or higher hydrodynamic modes, the effects of vortices dominate and the drive mechanism must be explained by the discrete-vortex model.

The jet-drive model concentrates on the jet-edge interaction, which induces a pressure difference Δp across the mouth. This pressure difference changes its sign each period of the tone produced by the flue instrument and sustains the acoustic wave in the pipe. It was first proposed by Coltman (1976) and calculated by Verge *et al.* (1994) with the help of conformal mapping technique. Incidentally, the results of Verge *et al.* corresponded to the source description of the edge tone by Powell (1961), who used Curle's equation to calculate the far field of the sound produced by an edge tone configuration. The acoustic field is dominated by a dipole field corresponding to the force

$$Force = \int p' \vec{n} dS \,. \tag{6.4}$$

This is the force exerted by the flow on the labium. According to Newton's third law, the lumped air at the mouth hole is driven by a force supplied by the labium (Fig. 6.19).

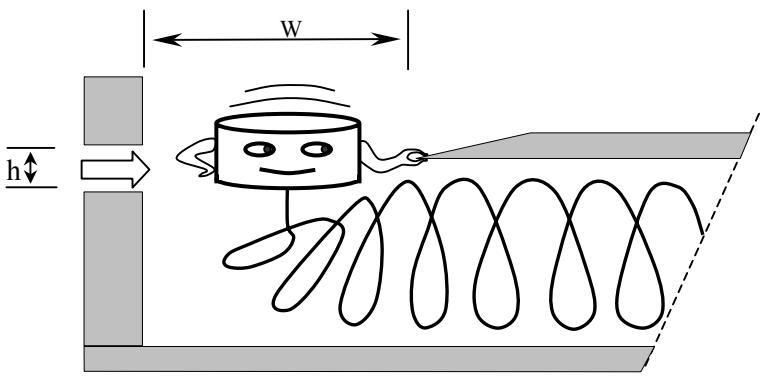


Figure 6.19: *Schematical depiction of the jet-drive model of a recorder-like flue instrument.*

In the jet-drive model, the jet flow Q_j is split by the labium into the flow Q_{in} injecting into the resonator, and the flow Q_{out} not entering the resonator. According to conservation of mass, it is obvious that

$$Q_j = Q_{in}(t) + Q_{out}(t)\,. \tag{6.5}$$

Q_{in} and Q_{out} can be written as

$$Q_{in}(t) = \frac{Q_j}{2} + Q_{in}'(t) \tag{6.6}$$

$$Q_{out}(t) = \frac{Q_j}{2} + Q_{out}'(t) \tag{6.7}$$

with

$$Q_{in}'(t) = -Q_{out}'(t)\,. \tag{6.8}$$

The two complementary monopole volume sources Q_{in}' and Q_{out}' are depicted in Fig. 6.20(a). The streamlines of Q_{in}' are shown as dash lines in Fig. 6.20(b)(d). These two monopoles form a dipole that has the character of a force, as Rayleigh has predicted. In a lumped model, these two monopole sources are assumed to be placed on either side of the edge with a definite distance, shown as Q_1 and Q_2 in Fig. 6.20(b).

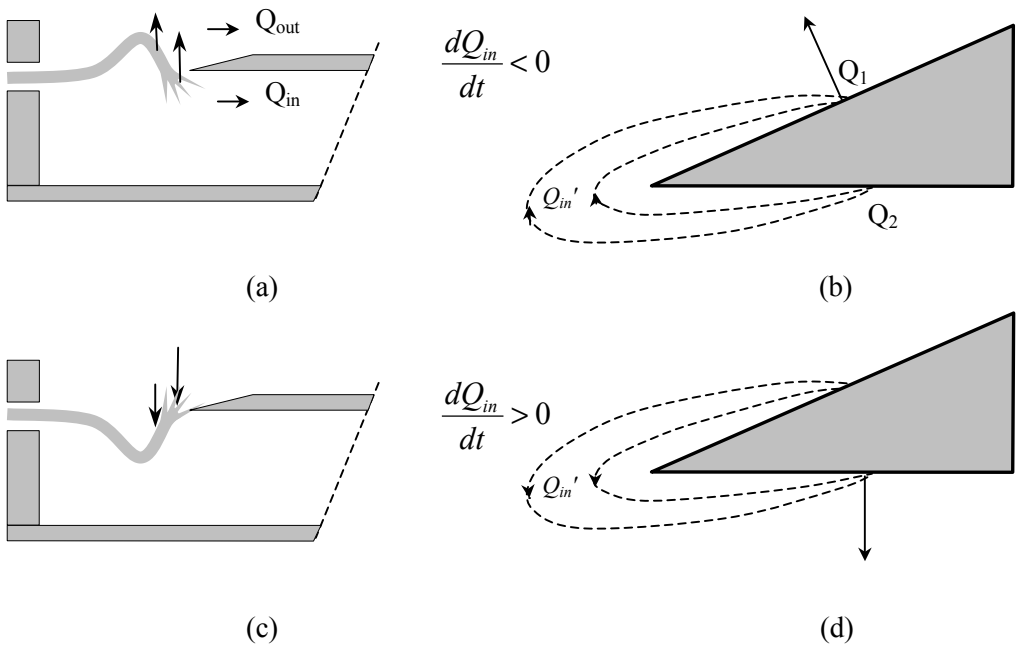


Figure 6.20: *Pressure difference induced by the time deviation of the injection into the pipe (Q_{in}) . The points Q_1 and Q_2 present two complementary monopole volume sources. The arrows in (b) and (d) represent the force driving the air at the mouth hole supplied by the labium.*

In the jet-drive model, these two complementary monopole sources induce a pressure difference Δp across the mouth. This pressure difference, driving the air at the mouth hole, is in proportion to the time deviation of Q_{in} (Verge 1994):

$$\Delta p = -\frac{\rho_0 \delta_d}{S_e}\frac{dQ_{in}}{dt}, \tag{6.9}$$

where S_e is the area of the mouth hole and δ_d is the "acoustic" distance between the monopole sources Q_1 and Q_2. Because the jet is observed to oscillate in phase of the acoustic flow at the mouth hole, the net force supplied by the labium has approximately the same direction of the acoustic flow. This implies a negative damping characteristic of self-sustained instruments.

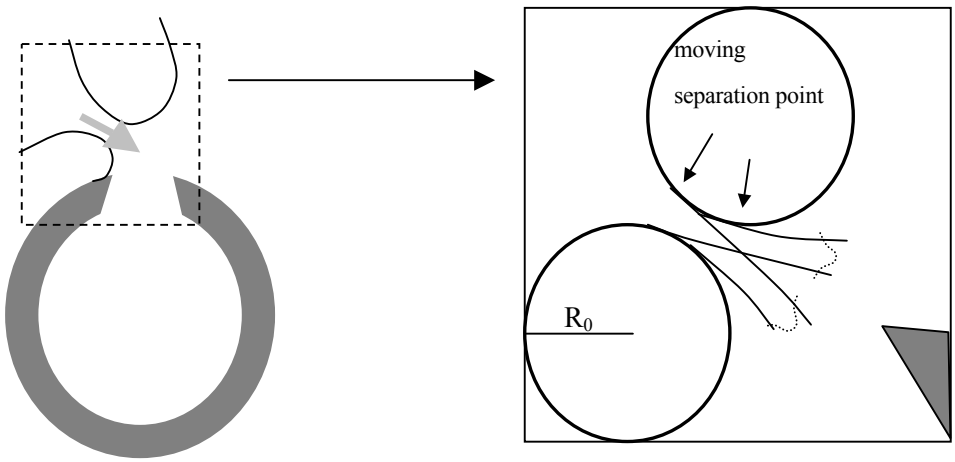
6.6.2 Jet's sensitivity to high harmonics

In the study of the Duffing oscillator model of the *dizi* membrane, the simulated results are in agreement with the tones generated by singing into the *dizi*, but there are considerable discrepancies between simulations and *dizi* tones. These discrepancies may be due to the assumption of a passive sound source. As this assumption is unrealistic, it is important to consider the responses of the vibrating jet at the mouth hole to high-frequency harmonics generated by the membrane. If the feedbacks of high frequencies are not negligible, the two non-linear mechanisms, the aerodynamics at the mouth hole and the membrane, will interact in a complicated manner.

The harmonics in the acoustic field at the mouth hole could affect the jet motion through several mechanisms. First, the instability of the jet appears to be the amplifier in a flue instrument. The jet responds to the transverse acoustic field, which is always dominated by the fundamental in a flue instrument. Many studies have been devoted to the question of whether the jet responds to the transverse flow velocity, transverse flow displacement or transverse flow acceleration (receptivity problem). Fabre *et al.* (2000) pointed out that both flow velocity and flow acceleration (pressure gradient) can be relevant. If flow separation point is fixed by a sharp edge (Fig. 6.19), there is no other distinction between velocity and acceleration drive than a $\pi/2$ phase lag. In the case of the flute, the lips of the player are rounded, and hence transverse flow acceleration can affect the position of the separation point (Fig. 6.21). For a given radius curvature R_0 of the edge, a significant contribution of the acceleration can be expected and the jet may be more sensitive to high-frequency harmonics, when fR_0/u' is of the order of unity or larger. For the harmonics around the first formant of *dizi* tones, $fR_0/u' \approx 2$.

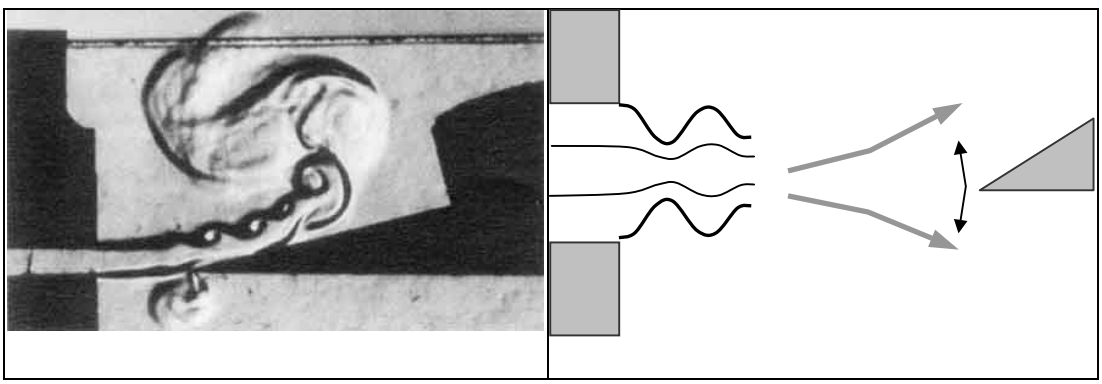
Another mechanism responsible for jet's sensitivity to high harmonics is its varicose oscillations due to the pressure fluctuations at the flue exit. Verge *et al.* (1994) showed that the pressure fluctuations at the mouth induce non-negligible jet velocity fluctuations at the flue exit. These volume flow fluctuations can induce varicose jet oscillations. This effect is more significant in flue instruments with shorter flue channel, because of the smaller inertia of the flue channel flow.

Varicose jet oscillation around 3–4 kHz observed by flow visualization was related to an edge-tone feedback loop (Verge *et al.* 1994). Ségoufin *et al.* (2000) also observed small vortices by flow visualization in the shear layers when the windway was short [Fig. 6.22(a)]. For a membrane flute, jet velocity fluctuations at the flue exit could be induced by the pressure fluctuations of the harmonics around 3–5 kHz generated by the membrane. According to Eq. (6.9), high-frequency varicose jet oscillations could result in rich harmonics in the pressure difference across the labium through the velocity fluctuations of the jet's shear layers [Fig. 6.22(b)]. This hypothesis should be examined by future experiments.



(a) (b)

Figure 6.21: *Effects of transverse flow acceleration on the separation points of the jet. (a) The embouchure of a flute. (b) The lips are approximated as cylinders with radius R_0 = 0.5 cm. The arrows represent the moving separation points of the oscillating jet, whose velocity profiles are presented by dash lines.*

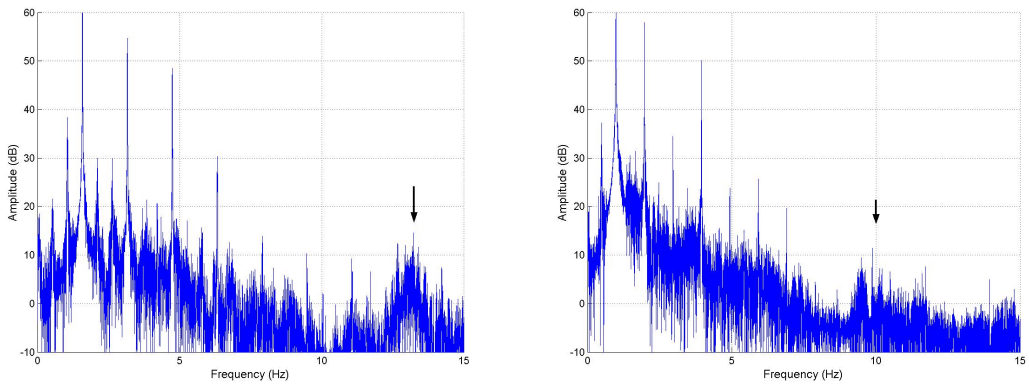


(a) (b)

Figure 6.22: *Varicose oscillations of the jet. (a) Small vortices in the shear layers of the jet. Flow visualization by Ségoufin et al. (2000), adapted with the author's permission. (b) Jet velocity fluctuations induced by the pressure fluctuations at the mouth.*

6.6.3 Transverse resonance of the *dizi* pipe

As the *dizi* membrane can generate high-frequency harmonics, it is important to consider the filtering effect of the transverse resonances of the *dizi* pipe. For membrane-less flutes, this filtering effect is only important for turbulence noises generation. Verge *et al.* (1997b) found that turbulence noises appear highly coupled to transverse resonances of the pipe. Fig. 6.23 compares the spectra of a *bangdi* tone and a large *dizi* tone in the fourth register. The frequency of the lowest transverse mode, estimated as $c/2D$ (D is the diameter), is indicated by the arrows. It can be observed the turbulence noises accumulate around the lowest transverse resonance of the pipe.



(a) (b)

Figure 6.23: *Comparison of the turbulence noises produced by a bangdi and a large dizi. (a) Tone spectrum of a bangdi with the inner diameter of 13 mm. The corresponding lowest transverse resonance is approximately 13.2 kHz. (b) Tone spectrum of a large dizi with the inner diameter of 17 mm. The corresponding lowest transverse resonance is approximately 10 kHz.*

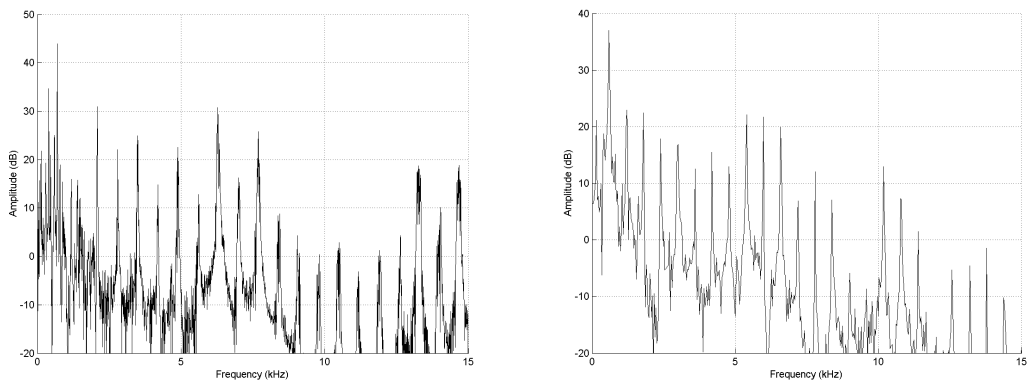


Figure 6.24: *Comparison of the spectra of a bangdi tone (left) and a large dizi tone (right). Note their second formant frequencies are approximately the same as the lowest transverse resonances shown in Fig. 6.23.*

For membrane flutes, the filtering effect of the transverse resonances plays a role in amplifying harmonics around the second formant. Fig. 6.24 displays the spectra of a *bangdi* tone and a large *dizi* tone with rich harmonics. Their first formants are both at 6 kHz, whereas their second formants have different center frequencies. The spectral envelops of harmonics (Fig. 6.24) and noises (Fig. 6.23) provide convergent evidence supporting a robust correlation between the second formant and the lowest transverse resonance of the pipe. It is noteworthy that while the first formant frequency increases with the tone intensity, the second formant frequency is invariant for a *dizi* and determined by the pipe diameter (see also Fig. 7.1).

6.7 Conclusions

The specific experiment on the *dizi* tone for which the membrane is located near a pressure node of the second harmonic wave shows the success of the quasi-sinusoidally driven Duffing oscillator model of the membrane in explaining the *dizi* harmonic generation in the range below 5 kHz. In principle, the present model is able to explain the three major spectral features of the *dizi* tones in the second and third registers: the first formant, the odd-numbered harmonics predominance, and subharmonics at high frequencies. Although this model fails to predict the spectral components above 10 kHz of *dizi* tones, its performance is good for soft *dizi* tones. For loud *dizi* tones, the acoustic field in the pipe is no longer dominated by low frequencies and the force driving the membrane contains high-frequency terms.

Rich harmonics above 10 kHz radiated by the membrane may be associated with the jet's sensitivity to the harmonics in the range 3–5 kHz generated by the membrane. If the jet amplifies these harmonics in the pipe, they may in turn drive the membrane to radiate harmonics above 10 kHz. The non-linear membrane, able to produce high harmonics, brings the question of the high-frequency response of the jet into prominence. This invites further physical studies on membrane flutes.

References

Coltman, J.W. (1968). Sounding mechanism of the flute and organ pipe, Journal of the Acoustical Society of America 44, 983–992.

Coltman, J.W. (1976). Jet drive mechanism in edge tones and organ pipes, Journal of the Acoustical Society of America 60, 725–733.

Dequand , S. (2000). Duct Aeroacoustics: from Technological Applications to the Flute. PhD thesis, Technische Universiteit Eidhoven, The Netherlands.

Elder, S.A. (1973). On the mechanism of sound production in organ pipes. Journal of the Acoustical Society of America 54, 1554-1564.

Fabre, B., Hirschberg, A., and Wijnands, A.P.J. (1996). Vortex shedding in steady oscillations of a

flue organ pipe. Acustica 82, 863-877.

Fabre, B., and Hirschberg, A. (2000). Physical modeling of flue instruments: A review of lumped models. Acustica 86, 599-610.

Fletcher, N.H. (1976). Sound production by organ flue pipes. Journal of the Acoustical Society of America 60, 926-936.

Fletcher, N.H., and Douglas L.M. (1980). The harmonic generation in organ pipes, recorders and flutes. Journal of the Acoustical Society of America 68, 767-771.

Fletcher, N.H., and Rossing, T.D. (1991). The Physics of Musical Instruments. Springer-Verlag.

Hirschberg, A., and Kergomard, J. (1995). Aerodynamics of wind instruments. In: Mechanics of Musical Instruments (edited by A. Hirschberg, J. Kergomard and G. Weinreich), Springer-Verlag, 291-369.

Howe, M.S. (1984). On the absorption of sound by turbulence and other hydrodynamic flows. IMA Journal of Applied Mathematics 32, 187-209.

Lighthill, M.S. (1954). On sound generated aerodynamically. Proceedings of the Royal Society of London, Series A 222, 1-32.

Meissner, M. (2002). Aerodynamically excited acoustic oscillations in cavity resonator exposed to an air jet. Acustica 88, 170-180.

Powell, A. (1964). Theory of vortex sound. Journal of the Acoustical Society of America 36, 177-195.

Verge, M.P., Caussé, R., Fabre, B., Hirschberg, A., Wijnands, A.P.J., and van Steenbergen, A. (1994a). Jet oscillations and jet drive in recorder-like instruments. Acustica 2, 403-419.

Verge, M.P., Fabre, B., Mahu, W.E.A., Hirschberg, A., van Hassel, R., Wijnands, A.P.J., de Vries, J.J., and Hogendoorn, C.J. (1994b). Jet formation and jet velocity fluctuations in a flue organ pipe. Journal of the Acoustical Society of America 95, 1119-1132.

Verge, M.P., Fabre, B., Hirschberg, A., and Wijnands, A.P.J. (1997a). Sound production in recorder-like instrument. I. Dimensionless amplitude of the internal acoustic field. Journal of the Acoustical Society of America 101, 2914-2924.

Verge, M.P., Hirschberg, A., and Caussé, R. (1997b). Sound production in recorder-like instrument. II. A simulation model. Journal of the Acoustical Society of America 101, 2925-2939.

Chapter 7.
Perception of harmonics I: Global spectral envelope

Abstract

The spectral distribution of harmonics of *dizi* tones can extend to 22 kHz, and their auditory attributes vary with the frequency region. The harmonics around the first formant allow the *dizi* to project through the orchestra, whereas the harmonics around and above the second formant can reduce breathiness by masking the turbulence noises.

The *dizi* tones in the fourth register are characterized by a poor spectral content. This inhomogeneity of brightness introduces interesting spatial effects in *dizi* music: the distancing effect and the localization-confusion effect. Since a sound loses more high frequencies while traveling through air, unexpected absence of high-frequency harmonics of *dizi* tones might be interpreted by the auditory system as an increasing distance from the *dizi*. Furthermore, a lack of spectral components above 4 kHz is likely to cause a confusion of sound elevation that is estimated according to high-frequency monaural cues. Although these two spatial effects could not actually change the *psychophysical* distance or elevation of the sound source, they appear to affect the *aesthetic* image of the *dizi* in music. The *dizi* with an inhomogeneous timbre is thus optimal for creating a multi-level auditory impression without changing the physical position of the instrument.

7.1 Introduction

The previous three chapters accounted for the physics of the *dizi*. In chapters 7–9 the focus turns to the issues concerning the perception of *dizi* tones. In exploring the perception of the harmonics of *dizi* tones, global spectral features are distinguished from local spectral features. This chapter concentrates on the global pattern of spectral envelopes of *dizi* tones, which is represented as the auditory attribute “brightness”. The local spectral feature, envelope jaggedness due to amplitude differences between flanking odd/even-numbered harmonics, will be discussed in chapter 8.

It is generally believed that musical timbres are mainly determined by their spectral envelopes

and temporal envelopes (for a recent review, see Risset and Wessel 1999). In all studies of the timbre spaces of musical instruments, brightness has appeared as the primary dimension responsible for the spectral envelope (Grey and Gordon 1978, Wessel 1979, Iverson and Krumhamsl 1993, Krimphoff *et al.* 1994, McAdam *et al.* 1995, McAdam and Winsberg 2000). In neuropsychological studies of timbre perception, brightness was also treated as the primary spectral parameter of sound stimuli, which was manipulated to test the effects of brain lesions on the sensitivity to timbre dimensions (Samson *et al.* 2002).

As the *dizi* has two formants with center frequencies at 4–6 kHz and 10–14 kHz, one-dimensional calculation models of brightness may be insufficient to describe the global envelope of its spectrum. Section 7.2 distinguishes the auditory attributes of the harmonics around/below the first formant and those around/above the second formant. In a concert hall, harmonics of *dizi* tones in different frequency regions have different radiation directionalities, which are discussed in section 7.3, related to *dizi* music recording.

The bright sound quality of *dizi* tones introduced by the membrane is not uniform through the tone range of the *dizi*. *Dizi* tones in the fourth register lose the rich spectral content characteristic of a membrane flute. The inhomogeneity of brightness induces some interesting auditory associations or illusions. Lin (1997) mentioned the spatial effects of abrupt brightness changes in *dizi* music, i.e., when a melody ascends from bright notes to dull notes, the sound source seems to move to a remote place and become more difficult to localize. Section 7.4 gives the psychoacoustic basis of these spatial effects. Musical examples of brightness changes in *dizi* tones are analyzed in section 7.5.

7.2 Frequency regions

7.2.1 Auditory brightness: Definition and biological implications

The auditory attribute "brightness" is related to the proportion of the energy distributed on high frequencies. Many semantic labels, such as "density" (Guirao and Stevens 1964), "sharpness" (von Bismarck 1974) and "nasality" (Kendall and Carterette 1993), all fall under the umbrella term "brightness". The diversity of these semantic labels associated with the spectral energy distribution seems to reflect that the sensation evoked by high-frequency components is multifaceted as well as context-dependent. Moreover, clusters of spectral components at different frequency regions may be responsible for different auditory sensations. However, traditionally brightness has been treated as a one-dimensional auditory attribute in psychoacoustic models and timbre spaces. A prevailing calculation model of brightness in musical timbre research is the *spectral centroid*, defined as

$$\text{Spectral centroid} = \sum_k kA_k \Big/ \sum_k A_k \tag{7.1}$$

where k is the rank of the harmonic and A_k is its amplitude.

As the semantic label implies, bright sounds are likely to arouse exciting auditory pictures. Bright *dizi* tones are often used to depict sunny scenes or highlight a mental state change in theatre. In some melodramas or films, beautiful Chinese landscapes are accompanied by bright *dizi* melodies. In Taiwanese *Gua* opera, it has become a convention that the excited mental state of falling-in-love-at-first-sight is represented by *dizi* tones.

The fact that high frequencies tend to evoke exciting auditory sensations can be dealt with in physiological, phylogenetic and ecological terms. From a physiological view, bright sounds are more exciting than dull sounds, partially because more auditory neurons responsible for high frequencies are excited. Auditory neurons with high characteristic frequencies (above 10 kHz) are excited only by loud, bright sounds. An analogue can be found in visual sensation, where the cone retinal is excited only by bright lights.

Regarding the rise of the new field of "biomusicology" (Wallin 1991), which places the analysis of music origin and its application to the study of human origins at its very foundation (Brown *et al.* 2001), the role of high frequencies in music can also be discussed from an evolutionary or phylogenetic viewpoint. High-frequency hearing is a characteristic unique to mammals, as neither fish, amphibians, reptiles, nor birds are able to hear above 10 kHz (e.g., Carr 1992, Köppl and Manley 1992). Masterton *et al.* (1968) attributed this ability of mammals to the ossicular linkage in the middle ear and proposed that mammals evolved this for horizontal sound localization by means of high-frequency binaural cues. Heffner and Heffner *et al.* (1992) added that the pinna of mammals was developed to resolve front-back confusion and judge the elevation of sound sources by providing high-frequency spectral cues (this issue will be discussed in detail in section 7.4.2). These evolutionary events in the middle/outer ear initiated improvements in cochlear mechanisms to enable mammals to perceive high frequencies (Manley 2000).

The ecological role of high frequencies in sound source localization provides new perspectives on the sensation of brightness. Besides sound source direction, high frequencies also contain cues for sound source distance. Strong high-frequency components could mean that the sound source is nearby, because high frequencies attenuate more rapidly than low frequencies when traveling through air (this issue will be further discussed in section 7.4.1). Consequently, as soon as humans or animals hear a sound with rich high-frequency components, they are likely to be excited, even alarmed. The auditory system alerts them to the presence of prey in the dark and around corners, or an approaching danger. This partially explains why high frequencies can be pleasant, in which case the listener is excited for prey; or unpleasant, for instance the listener is alarmed by danger. In the latter case the high frequencies may be related to the auditory attribute *sharpness* instead of brightness.

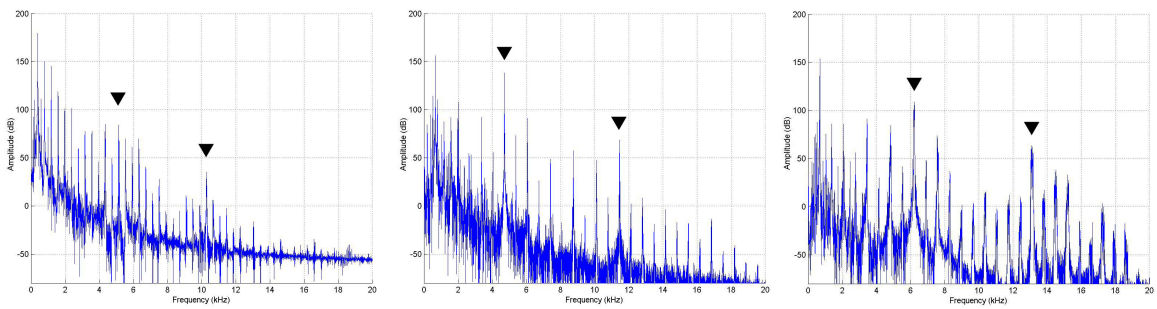
While brightness is associated with pleasant and exciting auditory sensations, sharpness often borders on annoyance. Bismarck (1974) studied the auditory attribute evoked by high frequencies in terms of sharpness, an antonym of sensory pleasantness. It is noteworthy that calculation models of

sharpness take into account a frequency-dependent weighting factor. In Zwicker and Fastl's (1999) model, the weighting factor is +1 for 1–16 bark and then increases from +1 to +4 for 16–24 Bark (4–20 kHz). Although this model ignores the fact that spectral components at low and mid frequencies can also evoke pleasantness, it is clear that spectral components above 10 kHz are particularly responsible for sharpness, as their sharpness weighting factors are greater than +3. This implies that the harmonics of *dizi* tones around the first formant and those around the second formant are likely to have different auditory attributes.

7.2.2 Center frequencies of the two formants

For a number of musical instruments, it is sufficient to model brightness as a one-dimensional auditory attribute, since spectral components of their tones are distributed at low/mid frequencies and only one formant is of importance. In the case of the *dizi*, the center frequencies of its two formants are separated for about 5 kHz. Regarding the wide spectral distribution of *dizi* tones, the traditional one-dimensional model of auditory brightness may oversimplify *dizi* timbre perception.

As have discussed in chapter 6, the center frequencies of the first formant of the *dizi* tend to increase as the tone intensity increases, whereas the center frequency of the second is mainly determined by the pipe diameter. Fig 7.1 compares the spectra of three soft *dizi* tones produced by professional musicians with *dizi* of different sizes. As can be seen, f_1 ranges from 4 to 6 kHz and f_2 increases from 10 to 13 kHz as the *dizi* size decreases.



(a) (b) (c)

Figure 7.1: *Formant frequencies of the dizi. (a) Soft tone, large dizi in G. (b) Soft tone, qudi in C. (c) Soft tone, bangdi in F.*

7.2.3 First formant: Stream segregation and loudness

The harmonics around the first formant of the *dizi* allow its sounds be easily distinguished from other instruments in *Kunqu* opera, thus reinforcing the leading role of the *dizi* in theatre. Because the membrane brings a bright timbre to the flute, *dizi* tones can be clearly heard by all performers and audiences, even when the singing and the percussion instruments are fairly loud. This is consistent with a general belief that brightness can best articulate stream segregation (Wessel 1979, McAdam and Bregman 1979). Because of the spectral energy accumulating around f_1, the *dizi* is able to project

on to the whole orchestra, as can be observed in many concerto-like *dizi* works composed in the past decades.

When the *dizi* is not a solo instrument, but played in a modern Chinese orchestra, it often has a relatively taut membrane, so that its sounds are not too overpowering. This principle of membrane-sealing also holds for the Korean membrane flute, the *taegum*. In ensemble groups, the *taegum* membrane is partially covered by a plate to reduce the coupling between the membrane and the air column in the pipe in order to produce a mellow timbre. By contrast, the membrane is mostly exposed in solo pieces to produce intense buzzing sounds (Howard 1988).

The first formant of *dizi* tones plays a key role in determining loudness. Equal loudness contours in the hearing range show a dip in the frequency range between 3 and 5 kHz (e.g., Zwicker and Fastl 1990). The *dizi* membrane effectively raises loudness of flute tones by transfering acoustic energy from lower frequencies to higher frequencies, to which our auditory system is more sensitive.

7.2.4 Second formant: Sharpness and breathiness

The harmonics around the second formant of a *dizi* tone are responsible for auditory sharpness. In Zwicker and Fastl's (1990) model of sharpness, the weighting factor for the two highest critical bands (10–20 kHz) is approximately three times greater than that below 16 Bark (< 4 kHz). As the *dizi* tones in the fifth register always have strong components above 10 kHz, these components in the two highest critical bands can give a sharp sound quality to these *dizi* tones.

Although sharpness is regarded by psychoacousticians as a synonym of auditory unpleasantness, musical instruments able to produce strong components above 10 kHz are widely used. For instance, the triangle sounds have strong spectral components above 10 kHz (Fletcher and Rossing 1991). In orchestrations, its timbre is known as shimmering and exciting instead of unpleasant. The *dizi* tones in the fifth register also have such a shimmering timbre. From a musical view, it remains to be clarified why sharp sounds, presumed to be unpleasant, are often appreciated a great deal in music.

The second formant of the *dizi* plays an important role in reducing breathiness evoked by turbulent noises. It is mentioned in sections 6.2.1 that turbulence noises accumulate at the lowest transverse resonance of the pipe, typically in the frequency region between 10 and 14 kHz. In bright *dizi* tones, these noises are partially masked by the harmonics around the second formant. But for the *dizi* tones in the fourth register, no harmonics around the second formant are generated and hence these noises are not masked. This explains why the timbre of the *dizi* tones in the fourth register is not only characterized by low brightness, but also by high breathiness. In phoniatrics, breathiness of pathological human voices is also associated with turbulent noises and a relatively low intensity of harmonics (e.g., de Krom 1995).

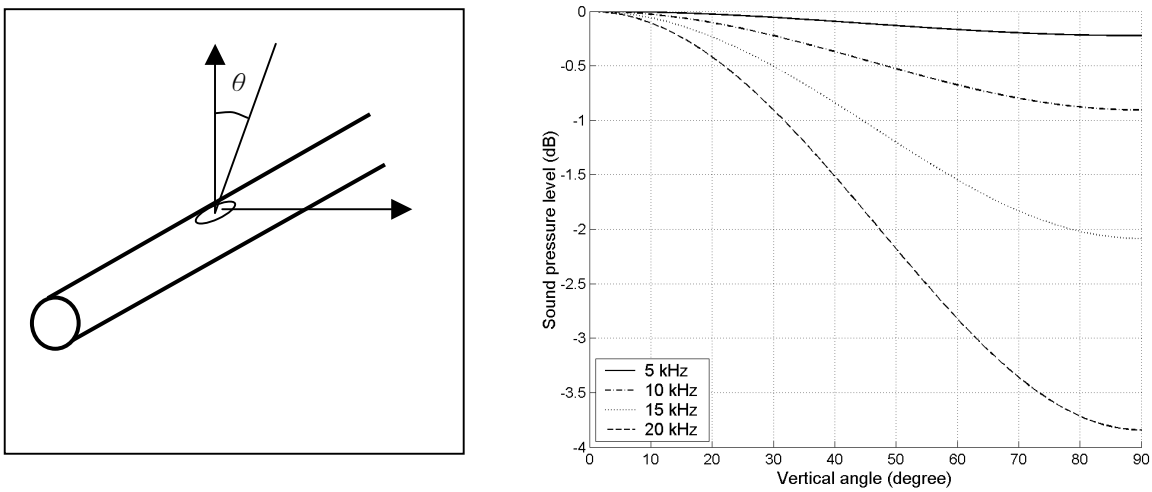
7.3 Radiation directionality

As the membrane of the *dizi* is a piston-like radiator, its radiation is not isotropic for spectral

components with wavelengths comparable with the membrane's diameter. The harmonics around the second formant of the *dizi* are more concentrated in the vertical direction than those around the first formant. Their radiation directionalities are studied here by taking the membrane to be a flanged plane source. The acoustic pressure p for the far field is

$$p \approx \frac{1}{2} j\omega\rho u a^2 \left(\frac{e^{-jkr}}{r} \right) \left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta} \right] \tag{7.2}$$

where θ is the vertical angle defined by Fig. 7.2(a); J_1 is a Bessel function of order one; and $k = 2\pi f/c$. The factor in the square bracket in Eq. (7.2) is a function of θ, determining the directionality for the frequency f. This factor has been plotted for four frequencies of 5, 10, 15 and 20 kHz in Fig. 7.2(b).



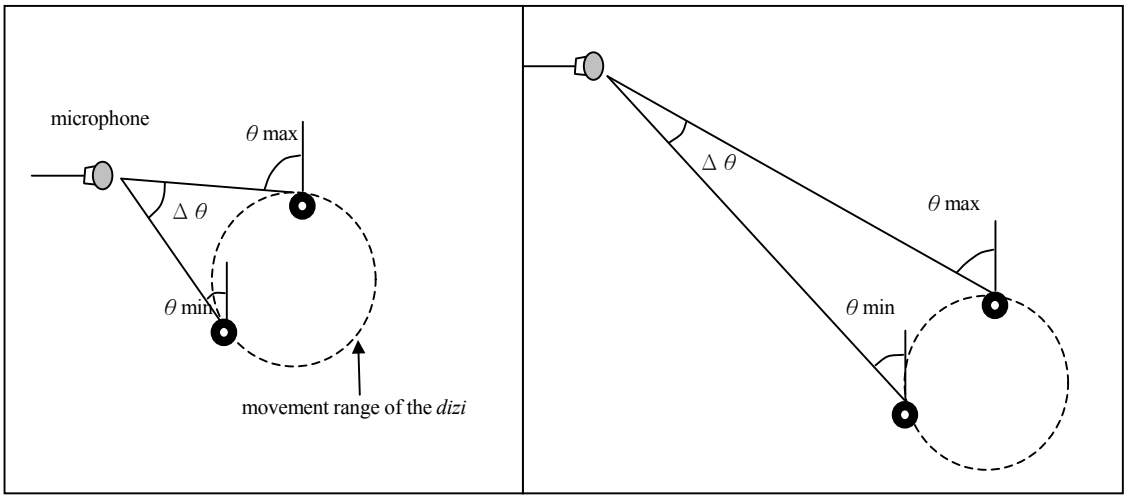
(a) (b)

Figure 7.2: *Radiation directionality of the dizi membrane. (a) Angular frame of reference of the dizi with its membrane pointing upward. (b) Relative magnitude distribution along θ for 5, 10, 15, and 20 kHz. These curves are plotted according to the bracket in Eq. (7.2) with the membrane radius a = 5 mm.*

In comparison with the horn mouth of a brass instrument or the human mouth, the size of the membrane is too small to result in significant radiation directionality for frequencies below 10 kHz. However, the radiation directionality of the harmonics above 10 kHz cannot be ignored and plays an important role in timbre perception. Although the *dizi* is able to generate harmonics with frequencies higher than 10 kHz, only a fraction of them directly reaches listeners. When $\theta = 90^0$, corresponding to the normal position of listeners, the sound level difference between 5 kHz and 18 kHz is 3 dB. Consequently, sharpness induced by the harmonics around and above the second formant is reduced

by their radiation directionality. In a concert hall, listeners at higher levels are likely to receive stronger high frequency harmonics of the *dizi*.

The radiation directionality of high-frequency harmonics is especially important for recording engineers. If *dizi* tones are recorded with a microphone located near the musician, the spectral content will vary with movements of the instrument because the radiation of high frequencies is concentrated in the direction the membrane is pointing. Fig. 7.3 shows $\Delta\theta$ corresponding to a movement range of the *dizi*. As the curves in Fig. 7.3(b) show, a larger $\Delta\theta$ leads to more amplitude fluctuation of the high-frequency harmonics, thus resulting in an unnaturally varying sound quality in the recorded *dizi* tones. This problem, noticeable in some live recordings of *dizi* music, can be removed by positioning the microphone a greater distance from the musician. It is suggested from Fig. 7.3(b) that the distance between the microphone and the *dizi* membrane should be three or four times larger than the movement range of the musician during performance.



(a) (b)

Figure 7.3: *Importance of the distance between the microphone and the dizi with respect to the radiation directionality of high-frequency harmonics. (a) Microphone located near the musician. As* $\Delta\theta > 50^0$*, the dizi movements will cause remarkable amplitude fluctuation of high-frequency harmonics in recording. (b) Microphone located more distant from the musician. As* $\Delta\theta < 30^0$*, the dizi movements will cause less amplitude fluctuation of high-frequency harmonics in recording.*

7.4 Spatial effects induced by brightness changes

As the *dizi* membrane is driven by acoustic pressure, it will hardly vibrate when it is located at or near a pressure node of the fundamental wave. It is well known by *dizi* players that the *dizi* loses its bright timbre for the notes in the fourth register. This brightness inhomogeniety is intrinsic for all membrane flue instruments, but it is treated very differently in China and Korea. Most contemporary *dizi* players in China dislike the brightness inhomogeneity. They believe that the timbre of a good

musical instrument should be uniform. Yet, Lin (1997) indicated that the varying timbre of the *dizi* is optimal for introducing interesting spatial effects into music and for enriching the tone colors of *dizi* melodies. A similar belief prevails in the circle of Korean traditional musicians. They state that the variety of tone colors across various registers makes the *taegum* an ideal solo instrument (Howard 1988).

One of the auditory associations elicited by brightness changes is that the *dizi* sounds like a human voice switching from the modal to the falsetto register when a melody ascends from the third to the fourth register. The falsetto produces weaker high-freuqnecy harmonics than the modal register produced (Large *et al.* 1972, Monsen and Engebretson 1977). In the other words, the falsetto voice source is more dominated by the lower source-spectrum harmonics. The same spectral feature can be found in the *dizi* tones in the fourth register. It is interesting to note that switching between the modal and falsetto registers is a very important singing technique for young male roles in *Kunqu* opera.

When the melody ascends from the third to the fourth register of the *dizi*, a dramatic decrease in brightness evokes some interesting auditory effects or illusions: the sound source becomes remote, and difficult to localize. These spatial effects have been deliberately utilized by *dizi* composers. For instance, the composer and performer Yu xunfa successfully used various *dizi* tone colors in *Langye Shenyun* (A Wonderful Melody of Mountain *Langye*) to create a multi-leveled auditory image. By switching from the second playing mode to the third playing mode with the same fingering, the *dizi* jumps from the third to the fourth register. It sounds as if the instrument moves to a remote place. This effect will be called the 'distancing effect' of *dizi* tones in the present thesis. Moreover, this dull tone is especially difficult to localize in comparison to normal, bright *dizi* tones. These two effects, the distancing effect and the localization-confusion effect, are rather striking in a concert hall, as the audience can see the player standing in the same place on stage but hear an illusion of sound source location change.

Although these spatial effects in *dizi* music have been experienced by audiences and mentioned by musical aestheticians (e.g., Lin 1997), their theoretical basis remains untouched. Here the spatial effects will be studied theoretically in terms of the perception of auditory distance and sound source localization.

7.4.1 Distancing effect and sound absorption of air

A change in physical distance between a sound source and a listener produces a variety of changes in the acoustic properties of the sound reaching the listener. The principal acoustic factors or cues for auditory distance perception include intensity, direct-to-reverberant energy ratio, spectrum, and binaural differences (e.g., Zahorik 2002). Even non-acoustic factors such as vision can contribute to perceived distance. In *dizi* music, the distancing effect is, in principle, a result of a dramatic change of spectral envelope.

The physical foundation of the spectral envelope change due to an increasing distance from the sound source is the sound attenuation in air. Sounds radiated from a remote source lose more high

frequencies while traveling through air, whereas there are fewer energy losses for low frequencies. The mechanisms of air absorption include viscosity, thermal conduction, and energy interchange between molecules with differing external excitation levels (Bass *et al.* 1990). Bass *et al.* (1995) proposed a formula for calculating the sound absorption coefficient (in decibels per 100 meters), which is plotted in Fig. 7.4 for various values of relative humidity.

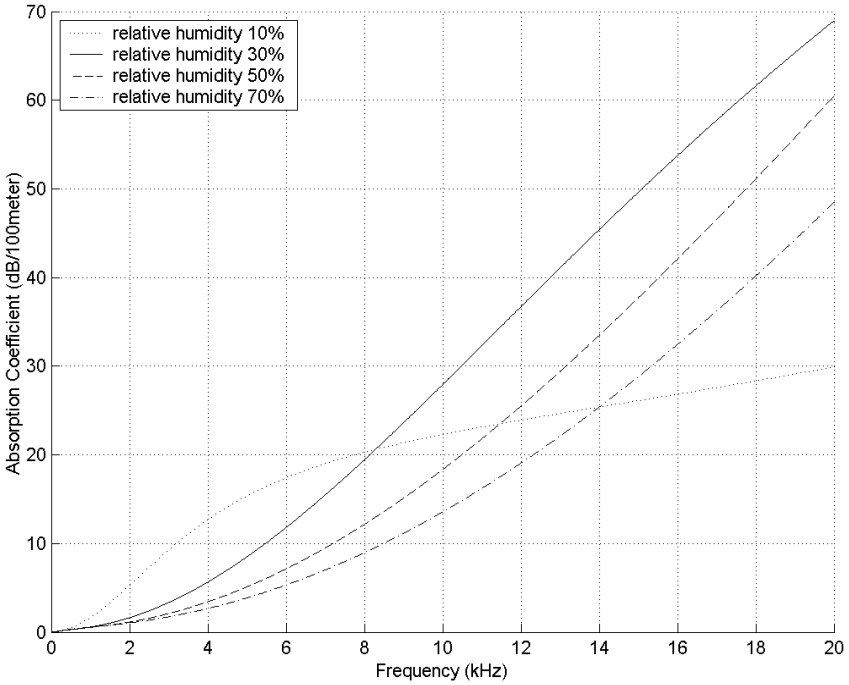


Figure 7.4: *Sound absorption coefficient (in decibels per 100 meters) for air at 20°C and 1 atmosphere calculated according to Eq. (3) in Bass et al. (1995).*

As can be seen, the sound absorption coefficient is an increasing function of frequency. Air absorption can be disregarded at low frequencies, but it becomes important for frequencies higher than 4 kHz. At 10 kHz, the sound attenuates 10–30 dB per 100 meters. Therefore, remote sounds are generally characterized by dominant low and mid frequencies.

With the sound absorption coefficient, it is easy to estimate the "physical" distance in the distancing effect that occurs when a *dizi* melody ascends from lower registers to the fourth register. If we start with a *dizi* tone in the third register, its spectrum at an arbitrary distance can be calculated. Fig. 7.5 compares the spectra of such a bright *dizi* tone at distances of 1, 100 and 500 m. As the distance increases, upper harmonics attenuate faster than lower harmonics. At the distance of 500 m, the spectral envelope shows a slope approximately 10 dB/kHz. The spectrum of a *dizi* tone in the fourth register has a similar spectral envelope slope. Therefore, if the auditory system assumes a constant spectral envelope of the *dizi* melody, a dramatic decrease in brightness in a *dizi* melody might be interpreted as if the sound source had moved to several hundred meters away.

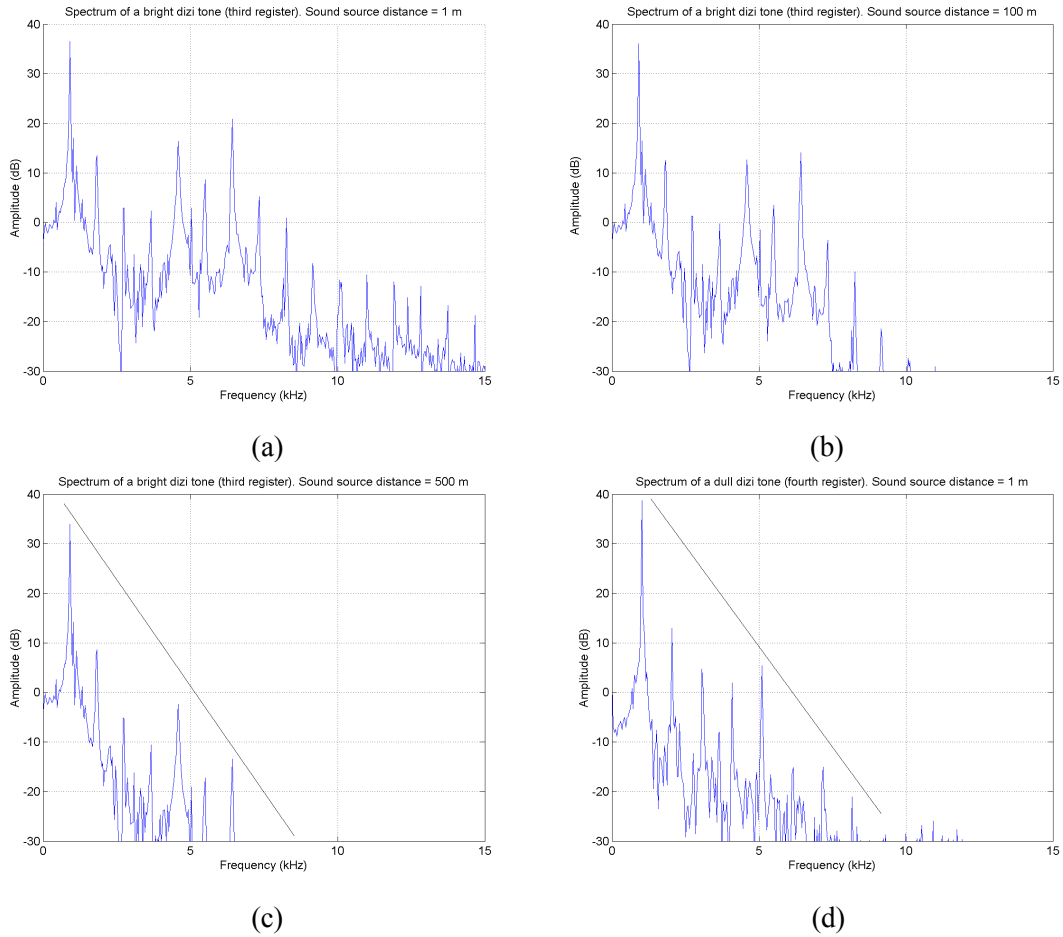


Figure 7.5: *Distancing effect in dizi music and its relation to air absorption. Spectrum changes in (a)(b)(c) due to distance increases have been calculated for relative humidity 50%. The dizi tones in (a) and (d) are excerpts from "Kongshan Wangyue". (a) Spectrum of a tone in the third register of the large dizi at the distance of about 1 m. (b) Spectrum of the same tone at the distance of 100 m. (c) Spectrum of the same tone at the distance of 500 m. (d) Spectrum of a tone in the fourth register at the distance of 1 m. Note the similarity between this spectral slope and that of the tone in the third register at 500 m.*

One of the psychological foundations of the distancing effect is the relatively constant brightness of a sound object. In the natural world, many situations exist in which sensory input is perceived to be constant in the presence of profound changes in the physical stimulus affecting the sensory receptors. For example, Zahorik and Wightman (2001) showed a robust loudness constancy with changing sound source distance, similar in many ways to visual size constancy with changing object distance. The majority of musical instruments show spectral shape constancy with varying pitch. The distancing effect of the *dizi* is based on the assumption of brightness constancy. As the auditory system has adapted to *dizi* tones with a high value of brightness, it will be puzzled by a dramatic decrease in brightness and may associate it with an increasing distance from the sound source.

Although the distancing effect in *dizi* music can be explained in terms of air absorption, it is important to note that the perception of sound distance is a complicated and flexible cognitive process, where many acoustic cues are combined and weighted for judgments of sound distance (e.g., Wightman and Kistler 1997, Zahorik 2002). Next to the spectral cues, sound intensity is also a robust acoustic distance cue. If a decrease in brightness is accompanied by a decrease in sound intensity, these two coherent cues are likely to give an auditory picture of a distancing sound source. Conversely, if the spectral receding cue is inconsistent with other cues, the perception of sound distance could be ambiguous.

A decrease in brightness in *dizi* music may not actually change the "psychophysical" distance of the sound source. Instead, it may affect the "aesthetic" image of the *dizi* in music. In various pieces for the *dizi*, the spectral distance cue interacts with other auditory attributes such as loudness and pitch. From this viewpoint, it should be more important to analyze the general attribute of the distancing effect in music rather than to treat it as an isolated acoustic phenomenon. Musical examples are given in section 7.5 to demonstrate the role of the distancing effect in various musical contexts.

7.4.2 Localization-confusion and monaural spectral cues

For the dull notes in the fourth register of the *dizi*, the absence of harmonics above 5 kHz affects listeners' ability to locate the sound source. The musical impression of this difficulty in sound source localization is described by the Chinese word *piau-miau* （縹緲）. This word, roughly translated as "dim", is an important aesthetic aspect in traditional music, poetry and drawing. From an acoustic viewpoint, the effect of localization-confusion in dull *dizi* tones is related to the crucial role of high-frequency monaural spectral cues for sound source localization.

The horizontal localization of a sound source is believed to be processed with the interaural time difference (ITD) and the interaural level difference (ILD) (Blauert 1997). However, the auditory system cannot distinguish, based on these binaural cues, between all source positions with the same horizontal component that lie on the so-called "cone of confusion." (Blauert 1997, Wightman and Kistler 1992) There are essential features of human sound localization that cannot be explained by ITDs and ILDs. One obvious example is the vertical localization on the median plane, where ITDs and ILDs are presumed to decline to zero. Away from the median plane, the front-back symmetry also results in front-back confusion.

As mentioned in section 7.2.1, monaural spectral features due to pinna diffraction are the primary cues for resolving front-back confusion and estimate elevation of sound source. These cues result from the direction-dependent linear filtering of incoming sound waves accomplished by the pinna. Pinna effects start to appear at frequencies around 4 kHz, where the wavelength becomes comparable to the pinna size, with the so-called "pinna notch" appearing in the frequency region from 6 to 12 kHz (Shaw 1997). The cone of confusion is resolved if the sounds to be localized have components extending to higher frequencies. Carlie *et al.* (1999) compared localization estimates of all-pass, low-pass and high-pass stimuli filtered from a spectrally flat (400 Hz to 16 kHz) noise. The

removal of frequencies below 2 kHz had little effect on either the pattern of systematic errors or the distribution of localization estimates. In contrast, the removal of frequencies above 2 kHz resulted in a large increase in the cone of confusion errors.

In addition to the pinna effects, the low frequency cues due to head diffraction and torso reflections can also contribute to localization estimates. Algazi *et al.* (2001) used random noise bursts that were low-passed with the cut-off frequency of 3 kHz in the experiments of sound localization. They found that when the source was located away from the median plane, subjects were able to estimate its elevation, often with surprisingly good accuracy. The physical origin of the low-frequency features was attributed primarily to head diffraction and torso reflections.

While there is a great deal known about the manner in which the low-frequency and mid/high-frequency features are computed for sound source localization (e.g., Middlebrooks 1992, Jin *et al.* 2000, Algazi *et al.* 2001), less is known about how these cues are integrated in the auditory system. Moreover, the auditory system must recalibrate on a short time scale when a listener moves from one acoustic environment to another (Shinn-Cunningham 2000). From this view, the localization-confusion in *dizi* music may be explained by the limit of the short-term plasticity of the auditory system.

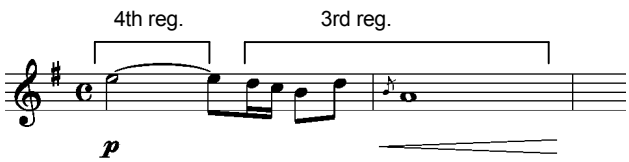
In flute music, the dull tones of Baroque flutes are not likely to produce localization confusion, whereas the dull *dizi* tones appear difficult to localize. When listening to *dizi* tones in the lowest three registers, the auditory system may adapt to localize the sound source using robust spectral cues above 4 kHz. When the spectral components in this frequency region are suddenly reduced or eliminated, the auditory system cannot immediately recalibrate to another mode in which the spectral cues below 4 kHz are also used for sound localization. Consequently, the dull *dizi* tones tend to cause more localization-confusion than the dull tones of Baroque flutes. When listening to Baroque flute music, the auditory system may rely on low-frequency cues for sound localization and no recalibration is needed.

Like the distancing effect, the localization-confusion effect of *dizi* tones in the fourth register interacts with other musical entities in various contexts. As the localization-confusion of dull tones mainly affect the listener's judgment of the elevation of the sound source, and due to the robust association between height and pitch, a decrease in brightness combined with an increase in pitch is likely to induce an auditory picture of the sound source becoming higher. When a *dizi* melody ascends from the third register to the fourth register, an increase of its "musical" height will be enhanced by elevation-confusion of the sound source due to a lack of spectral cues above 4 kHz. In certain musical contexts, this effect is combined with the distancing effect, giving an auditory picture of the sound flying higher and further away. The spatial effects in *dizi* music are thus optimal for creating a multi-level auditory impression without changing the physical position of the instrument.

7.5 Brightness in *dizi* music

7.5.1 Depiction of sunrise

An increase in brightness of *dizi* tones is used in the *qudi* solo piece "Morning" to depict the sunrise. In the first bar the melody descends from the fourth register to the third register. The second bar is a *crescendo* tone in the third register (Ex. 7.1).

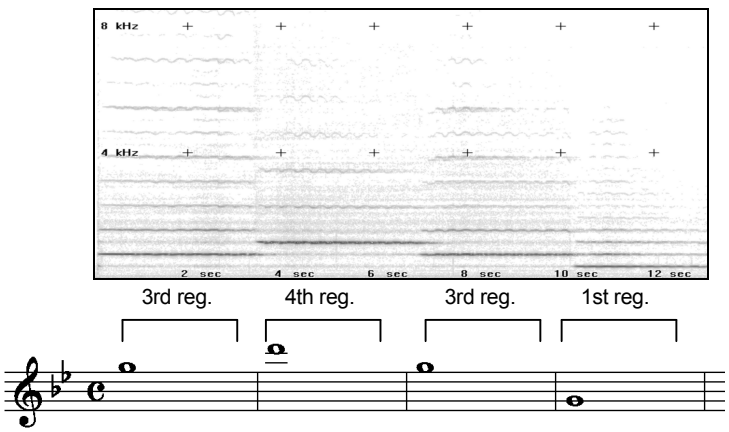


Example 7.1: *Opening bars of the qudi solo piece "Morning" composed by Zhao Songting.*

It is interesting to note that the sunrise is depicted by a *descending* melody rather than an ascending one, because the *dizi* tones in the third register are brighter than those in the fourth register. The brightness increase in the second bar results from an increase of tone intensity.

7.5.2 Spatial effects combined with vibrato

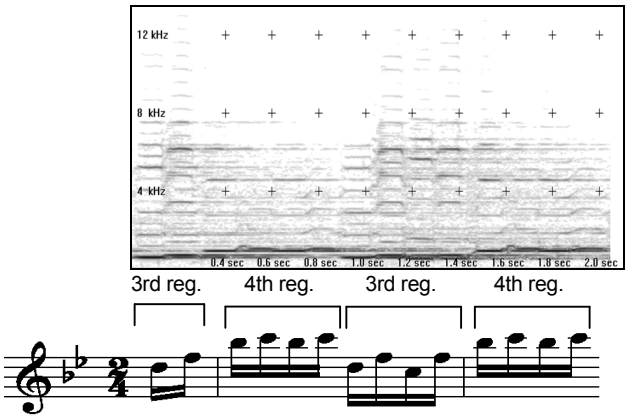
The *dizi* player and composer Yu Xunfa combined an abrupt change in brightness with vibrato in the large *dizi* solo piece "A Wonderful Melody of Mountain *Langye*". The four-tones melody is produced with the same fingering ●●●○○○○, going through the second, third, second, and first play modes (Ex. 7.2). As can be seen in the spectrogram, the second note in the fourth register has a spectral content poorer than the note in the third register. Hence, the second tone sounds particularly remote and dim because of the spatial effects due to brightness changes.



Example 7.2: *Spatial effects in the large dizi solo piece "A Wonderful Melody of Mountain Langye".*

7.5.3 Rapid changes in brightness

The *dizi* player and composer Wang Ciheng combined changes in brightness and pitch in the bass *dizi* solo piece "Kongshan Niaoyu" (Watching the Moon in an Empty Mountain). The melody moves in the third and fourth registers (Ex. 7.3). As lower notes in the third register are brighter than higher notes in the fourth register, the sound source seems to move up and down, and thereby this melody may call to mind a bird. In this example, decreases in brightness are combined with increases in pitch to give an auditory picture of the sound flying higher.



Example 7.3: *Spatial effects in the large dizi piece "Watching the Moon in an Empty Mountain".*

7.6 Conclusions

The auditory attributes of the harmonics of *dizi* tones vary with the frequency region. The harmonics around the first formant, responsible for brightness, contribute to articulating stream segregation of the *dizi*, whereas the harmonics around and above the second formant, responsible for sharpness, can reduce breathiness by masking the turbulence noises.

When a *dizi* melody moves form third or fifth registers to the fourth register, a dramatic decrease in brightness introduces two spatial effects: the distancing effect and the localization-confusion effect. Although these two spatial effects could not actually change the *psychophysical* distance or elevation of the sound source, they appear to affect the *aesthetic* image of the *dizi* in musical contexts, where the spatial effects are combined with other musical entities such as loudness, pitch and vibrato.

References

Algazi, V.R., Avendano, C., and Duda, R.O. (2001). Elevation localization and head-related transfer function analysis at low frequencies. Journal of the Acoustical Society of America 109(3), 1110-1122.

Bass, H.E., and Shields F.D. (1990). Absorption of sound in air: high-frequency measurements. Journal of the Acoustical Society of America 88, 2019-2021.

Bass, H.E., Sutherland, L.C., Zuckerwar, A.J. Blackstock, D.T., and Hester, D.M. (1995). Atmospheric absorption of sound: Further developments. Journal of the Acoustical Society of America 97(1), 680-683.

von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. Acustica 30, 160-171.

Brown, S., Merker, B., and Wallin, N.L. (2001). An introduction to Evolutionary Musicology. In: The Origins of Music (edited by N.L. Wallin, B. Merker and S. Brown), MIT Press, 3-24.

Carr, C.E. (1992). Evolution of the central auditory system in reptiles and birds. In: The Evolutionary Biology of Hearing (edited by D.B. Webster, R.R. Fay and A.N. Popper), Springer-Verlag, 511-544.

Carlie, S, Delancy, S., and Corderoy, A. (1999). The localization of spectrally restricted sounds by human listeners. Hearing Research 128, 175-189.

de Krom, G. (1995). Some correlates of pathological breathy and rough voice quality for different types of vowels fragments. Journal of Speech and Hearing Research 38, 794-811.

Fletcher, N.H., and Rossing, T.D. (1991). The Physics of Musical Instruments. Springer-Verlag.

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbre. Journal of the Acoustical Society of America 61, 1270-1277.

Grey, J.M., and Gordon, J.W. (1978). Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America, 63, 1493-1500.

Guirao, M., and Stevens, S.S. (1964). Measurement of auditory density. Journal of the Acoustical Society of America 36, 1176-1182.

Heffner, R.S., and Heffner, H.E. (1992). Evolution of sound localization in mammals. In: The Evolutionary Biology of Hearing (edited by D.B. Webster, R.R. Fay and A.N. Popper), Springer-Verlag, 691-716.

Howard, K. (1988). The *Taegum*. In: Korean Musical Instruments, Se-Kwang Music Publishing, 99-117.

Iverson, P., and Krumhansl C.L. (1993). Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America 94(5), 2595-2603.

Jin, C., Schenkel, M., and Carlile, S. (2000). Neural system identification model of human sound localization. Journal of the Acoustical Society of America 108(3), 1215-1235.

Kendall, R.A., and Carterette, E.C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. Von Bismarck's adjectives. Music Perception 10(4), 445-468.

Köppl, C., and Manley, G.A. (1992). Functional consequences of morphological trends in the evolution of lizard hearing organs. In: The Evolutionary Biology of Hearing, (Ed. by D.B. Webster, R.R. Fay and A.N. Popper) Springer-Verlag, 489-510.

Krimphoff, J., McAdams, S., and Winsberg S. (1994). Caractérisation du timbre des sons complexes.

II Analyses acoustiques et quantification psychophysique. Journal de Physique IV, Colloque C5, 4, 625-628.

Large, J., Iwata, S., and von Leden, H. (1972). The male operatic head register versus falsetto. Folia Phoniatrica 24, 19-29.

Lin, G.F. (1997). Dikuan Yioqin. （林谷芳《諦觀有情：中國音樂裡的人文世界》，台北：望月文化）

Manley, G.A. (2000). Cochlear mechanisms from a phylogenetic viewpoint. Proc. Nat. Acad. Sci. (USA) 97(22), 11736-11743.

Masterton, B., Heffner, H., and Ravizza, R. (1969). The evolution of high-frequency hearing. Journal of the Acoustical Society of America 45, 966-985.

McAdams, S., and Bregman, A.S. (1979). Hearing musical streams. Computer Music Journal 3(4), 26-43.

McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., and Krimphoff J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research 58, 177-192.

Middlebrooks, J. (1992). Narrow-band sound localization related to external ear acoustics. Journal of the Acoustical Society of America 92(5), 2607-2624.

Monsen, R.B., and Engebretson, A.M. (1977). Study of vibrations in the male and female glottal wave. Journal of the Acoustical Society of America 62, 981-993.

Samson, S., Zatorre, R.J. and Ramsay, J.O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. Brain 125, 511-523.

Shaw, E.A.G. (1997). Acoustical features of the human external ear. In: Binaural and Spatial Hearing in Real and Virtual Environments (edited by R. H. Gilkey and T. R. Anderson), 25–47.

Shinn-Cunningham, B. (2000). Learning reverberation: Implications for spatial auditory displays. International Conference on Auditory Display, 126-134.

Shinn-Cunningham, B. (2001). Models of plasticity in spatial auditory processing. Audiology and Neuro-Otology 6, 187-191.

Wallin, N.L. (1991). Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origins and Purposes of Music, Pendragon Press.

Wessel, D. (1979). Timbre space as musical control structure. Computer Music Journal 3(2), 45-52.

Wightman, F.L., and Kistler, D.J. (1992). The dominant role of low-frequency interaural time differences in sound localization. Journal of the Acoustical Society of America 91, 1648-1661.

Zahorik, P., and Wightman, P.L. (2001). Loudness constancy with varying sound source distance. Nature Neuroscience 4, 78-83.

Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. Journal of the Acoustical Society of America 111(4), 1832-1846.

Zwicker, E., and Fastl, H. (1990). Psychoacoustics: Facts and Models. Springer-Verlag.

Chapter 8.
Perception of harmonics II: Predominance of odd-numbered harmonics

Abstract

Dizi tones with weak even-numbered harmonics are characterized by the hollow and/or nasal sound qualities described by Helmholtz. Nasal voices produced with a water membrane in the nasal cavity have the spectral feature of a predominance of odd-numbered harmonics around 5 kHz and a sounding mechanism similar to those of the *dizi* tones.

An analogy is proposed, comparing the spectral feature of large amplitude differences between odd- and even-numbered harmonics and the visual pattern of luminance-contrasting gratings. As these gratings make a strong visual impression, so too does a predominance of odd-numbered harmonics evoke a distinct timbre. When this predominance occurs at high frequencies, interference between adjacent harmonics are largely reduced. This may explain why nasal *dizi* tones are described as 'sweet' by musicians.

The predominance of odd-numbered harmonics at high frequencies can induce additional pitches, which enrich the texture in *dizi* solo music. An autocorrelation model of pitch suggests the conditions for the existence of the additional pitches. Since these two conditions are always satisfied in the *dizi* tones in the second and third registers but not in most tones produced by other musical instruments, this multi-pitch phenomenon is fairly unique to the *dizi*.

8.1 Introduction

In chapter 7 the global spectral envelope was studied and related to the bright sound quality in *dizi* tones. This chapter accounts for the local spectral envelope feature of significant amplitude differences between the adjacent odd- and even-numbered harmonics. This type of spectral envelope jaggedness is distinguished from the spectral envelope jaggedness due to significant amplitude

differences between the adjacent harmonic/subharmonic components. The latter will be discussed in the next chapter.

In timbre research, local features of spectral envelopes have attracted less attention than the overall spectral envelope, which is related to auditory brightness. Spectral envelope jaggedness, sometimes called "irregularity", has seldom been regarded as a dimension in timbre spaces. The third dimension of Krumhansl's (1989) timbre space, originally labeled "spectral flux", was re-interpreted by Krimphoff *et al.* (1994) as spectral envelope irregularity, which was defined as the log of the standard deviation of component amplitudes from a global envelope derived from a running mean of the amplitudes of three adjacent harmonics. However, the correlation between the third dimension and the spectral envelope irregularity was medium (Krimphoff *et al.* 1994, McAdams *et al.* 1995). The predominance of odd-numbered harmonics in clarinet tones is frequently treated as "specificity" in timbre spaces (Krumhansl 1989, Winsberg and Carroll 1989b, McAdams *et al.* 1995), thus indicating that few musical instruments, at least amongst those taken in timbre research, produce tones with significantly jagged spectra. Another reason for the lack of attention paid to spectral envelope jaggedness may be that no auditory attributes have been found which correlate with this spectral feature.

From a physical point of view, jagged spectra of musical tones are most easily caused by a predominance of odd-numbered harmonics. In addition to the clarinet, tones with weak even-numbered harmonics at low frequencies can also be produced by stopped organ pipes and panpipes. Plucking a string at the midpoint, as in harp or sometimes in guitar playing, can also produce such tones. Among these instruments, however, only the clarinet and the harp have been taken into timbre spaces.

It is invalid that no auditory attributes have been related to jagged spectra. Hermann von Helmholtz (1877) described tones dominated by odd-numbered harmonics as hollow or nasal, although his observation have generally been neglected for more than a century. It is important to note that Helmholtz distinguished the auditory attribute correlated to the predominance of lower odd-harmonics (hollowness) from that of upper odd-harmonics (nasality). From his definition, the *dizi* appears to be the musical instrument most capable of producing nasal tones with odd-numbered harmonics predominating around in the range of 4 to 6 kHz, whereas the clarinet and panpipes are just able to produce hollow tones with the predominance of odd-numbered harmonics restricted to low frequencies. Helmholtz's hollowness and nasality are revisited in section 8.2 with new psychophysical evidence.

While the mechanics underlying the spectral feature of odd-numbered harmonics predominance was explored in chapter 6, this chapter accounts for the perception of this spectral feature. Section 8.3 focuses on the timbre perception of the odd-numbered harmonics predominance, whose importance is discussed from various aspects, including multi-dimensional timbre analysis, analogue to visual sensation and critical bandwidth.

In addition to the distinct timbre of *dizi* tones with a predominance of odd-numbered harmonics,

this spectral feature can result in a multi-pitch phenomenon when it is significant at high frequencies. Bright *dizi* tones in the second and third registers tend to have additional pitches moving parallel above the fundamental pitch contour. In section 8.4 this pitch effect is related to the shifted residue pitch evoked by equally spaced inharmonic components (Boer 1956, Schouten *et al.* 1962, Patterson 1973, Patterson and Wightman 1976). Pitch models in both the frequency-domain and time-domain are used in the preliminary research on the multi-pitch phenomenon of *dizi* tones with odd-numbered harmonics dominating at high frequencies.

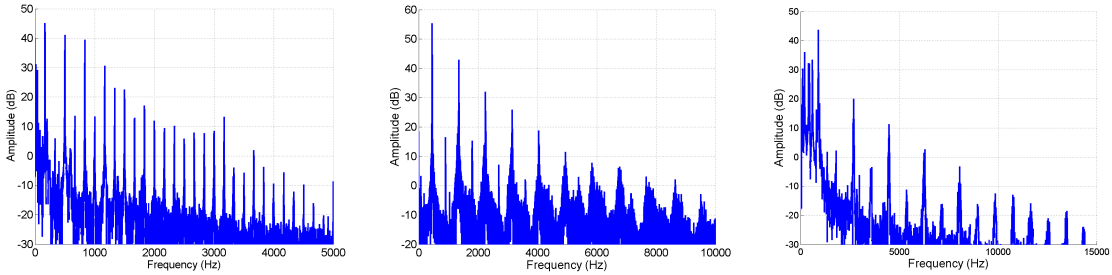
8.2 Helmholtz's hollowness and nasality revisited

The qualitative definition of the sound quality evoked by a predominance of odd-numbered harmonics in a complex tone was first given by Hermann von Helmholtz (1877). He reported that if only the odd-numbered harmonics were present, the quality of the tone was hollow, and when a large number of such upper harmonics were present it was nasal. It is interesting to note that Helmholtz's observations on these two sound qualities has attracted little attention from psychoacousticians or musicologists. The present study of the perception of *dizi* tones, predominated by odd-numbered harmonics, leads me to revisit Helmholtz's auditory hollowness and nasality.

8.2.1 Hollow sounds in music

While there are ambiguities of the definition of the hollow sound quality, some electrical sound engineers describe the sounds of square waves and triangle waves as "hollow and clarinet-like". The common feature shared by square waves, triangle waves and low clarinet tones is a predominance of odd-numbered harmonics.

The best-known musical tones characterized by a predominance of odd-numbered harmonics are low-pitched clarinet tones. Clarinet spectra show almost without exception an odd-harmonic predominance from the lowest tones through written top-line F, regardless of the pitch of the instrument itself (Gibson 1994). For low clarinet tones, the resonator is approximately an open-closed cylindrical pipe with the reed as the closed end. Therefore, only the odd-numbered harmonics can establish standing waves in an ideal stopped pipe, while the low even-numbered harmonics are suppressed. Fig. 8.1(a) shows the spectrum of a bass clarinet tone with significant amplitude differences between adjacent odd- and even-numbered harmonics.



(a) (b) (c)

Figure 8.1: *Spectra of hollow tones produced by woodwind instruments. (a) The bass clarinet. (b) The panpipe. (c) The recorder (accompanied by the organ).*

It is interesting to note that clarinet tones are seldom described as hollow,[2] which might be considered as a negative semantic label for Western musicians. The lowest register of the clarinet and the bass clarinet, usually called the *chalumeau register*, is described in orchestration textbooks as "warm and rich", sometimes "mysterious, shadowy or sinister" (Adler 2002). However, some composers were certainly aware of the hollow sound quality of these notes. For instance, Giuseppe Verdi (1813–1901) used unisons of two bass clarinets and two bassoons to create a "hollow and severe" sonority in his revised version of *Macbeth.*[3]

Besides reed instruments, flue instruments with stopped cylindrical pipes can also produce hollow sounds. Fig. 8.1(b) shows the spectrum of a panpipes tone, which was described as hollow because of the odd-numbered harmonics predominance (Fletcher and Rossing 1991). In spite of open cylindrical resonator, the flute and the recorder sometimes produce tones with the predominance of lower odd-numbered harmonics when the embouchure leads to fairly symmetrical jet vibrations [Fig. 8.1(c)].

The *dizi* produce sounds dominated by odd-numbered harmonics through a non-linear mechanism. When the membrane is driven by a sinusoidal acoustic pressure wave, its cubic non-linearity will generate odd-numbered harmonics (see chapter 6). *Dizi* tones in the second and third registers are thus perceptually comparable to those produced by stopped organ pipes or panpipes. Zhao Songting (1985) noticed that the bass *dizi* has a timbre similar to the clarinet. This timbre resemblance is likely to be due to the predominance of odd-numbered harmonics of their tones.

8.2.2 Psychological basis of hollow sound quality

In chapter 7 it was argued why the sounds containing strong high-frequency components are

---

[2] In a study of verbal attributes of the clarinet's timbre (Jost 1967), listeners were asked to describe clarinet notes within 60 attributes. But "hollow" was omitted from this verbal attribute list.

[3] Verdi's letter stated "The appearance of Hecate should be accompanied by the clarinet and bass clarinet in unison with the cello and bassoon. They will result in a hollow and severe sound (*suono cupo e severo come esige*), as the situation demands." (January 1865, letter to L. Escudier).

described as “bright”, a semantic label normally used in visual sensation. The same question can be asked in exploring auditory hollowness.

The label “hollowness” in audition can be explained in terms of an association between the shape of a sounding object and its sound spectrum; our cognitive system may have learned to judge whether an object is hollow according to its sounds. When a piece of hollow bamboo is exposed to a grazing wind flow along its opening, for instance, its “whistle” sounds will be dominated by odd-numbered harmonics. Such sounds may have been associated with hollow objects and labeled as hollow according to their shape through a learning process.

It remains to be clarified why the sounds produced by open cylindrical pipes or stopped conic pipes are not described as “hollow” in spite of their hollow shape. Note that their sounds do not possess spectral features that can easily be extracted for object identification. Hence, these hollow objects provide no distinct auditory cues worth learning for our cognitive system. On the other hand, the tones with dominant odd-numbered harmonics have a robust spectral feature for sound recognition. This feature differentiates the sounds produced by stopped cylindrical pipes from other sounds. We may have learned to extract this spectral feature and utilize it for judging the shape of objects emitting the sound, especially when they are not within the field of vision.

8.2.3 Nasality in speech and music

Although nasality in speech has been extensively studied for decades, its spectral cues are related to the overall spectral shape rather than significant amplitude differences between odd- and even-numbered harmonics, as proposed by Helmholtz (1877). It has been observed that the reduction in amplitude of the first formant is the primary cue of nasalization (e.g., Fant 1960). Nasal coupling during vowel production was also found to give rise to a pole and a zero in the region of the first formant introducing a spectral peak (House and Stevens 1956), whose frequency for nasalized vowels in German (Heike 1999), English and French (Chen 1997) has been reported. Another nasal peak between 250 and 450 Hz has been observed in studies of vowel nasalization (e.g., Dang and Honda 1995). The acoustic cues of hypernasality in children were also explored by formant analysis (Hawkins and Steven 1985).

In musical timbre research, Rusko (1997) systematically classified nasality of speech into four types: nasal consonants (with closed oral cavity), nasalized vowels, closed nose nasalization, and fully closed nasalization. Various nasal musical sounds were discussed within this classification. The clarinet, violin “con sordino”, trumpet and organ register “nasal”, all belonged to nasalized vowels, which were characterized by a weakening of the first formant and the emergence of additional formants. The predominance of odd-numbered harmonics of clarinet tones was not mentioned there.

8.2.4 Nasal voices produced by a membrane

The literature review cited above reveals that the nasal quality mentioned by Helmholtz (1877) has received no support in speech and music research; Helmholtz’s observation of nasality may be

either invalid, or imply a new type of nasal voice. I argue that Helmholtz's nasality corresponds to a new type of nasal voice, which has a sounding mechanism similar to the *dizi*.

The nasal voice correlated to Helmholtz's observation of nasality is produced when there is a water membrane in the nasal cavity. This membrane is presumed to have a cubic non-linearity and be driven by a quasi-sinusoidal acoustic wave generated by falsetto. Fig. 8.2 shows the spectrum of a voice produced by me with the mouth closed and recorded beneath the nose. I felt that a water membrane in the nasal cavity vibrated sympathetically while I was singing in falsetto register. This membrane produced rich odd-numbered harmonics. In the frequency region 2–5 kHz, a predominance of odd-numbered harmonics can be noticed. This preliminary experiment provides physical evidence for Helmholtz's observation that if odd-numbered harmonics dominate at high frequencies, the quality of the tone is nasal. A model for the production of this new type of nasality is schematically depicted in Fig. 8.3. It is noteworthy that the water membrane plays a role analogous to the *dizi* membrane in producing odd-numbered harmonics through a cubic non-linearity.

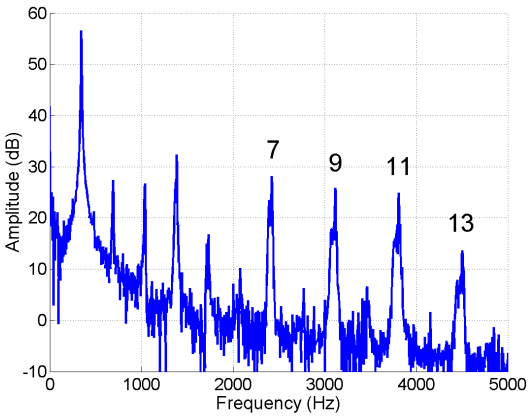


Figure 8.2: *Spectrum of a voice produced with the mouth closed and a membrane in the nasal cavity.*

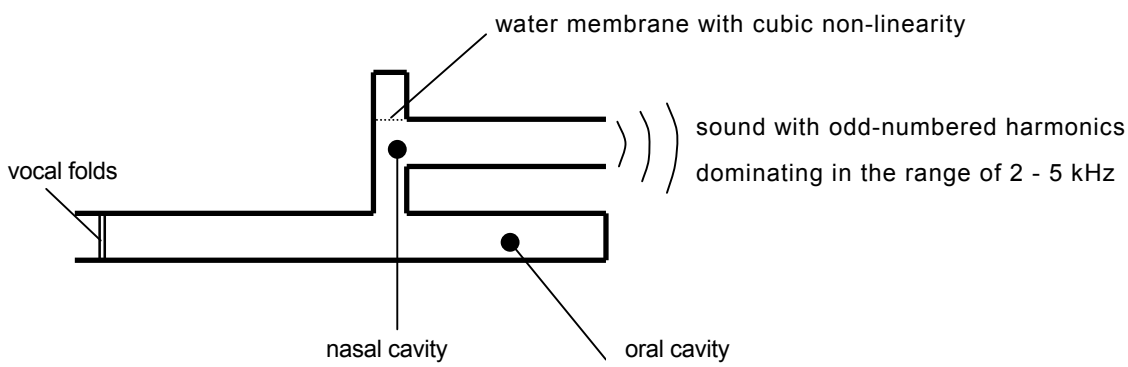


Figure 8.3: *Model of the production of nasal voices with the predominance of odd-numbered harmonics at high frequencies. A quasi-sinusoidal falsetto voice is generated by the vocal folds. If there is a water-membrane with cubic non-linearity in the nasal cavity, this membrane will transfer acoustic energy from the fundamental to higher frequencies by producing odd-numbered harmonics.*

8.3 Timbre perception

8.3.1 Odd-numbered harmonics predominance in timbre spaces

The above revisit of Helmholtz's auditory hollowness and nasality highlights the perceptual importance of the two sound qualities. However, in timbre research, the degree of odd-numbered harmonics predominance has never been treated as a dimension in timbre spaces, but a "specificity" of the clarinet (Krumhansl 1989, Winsberg and Carroll 1989b, McAdams *et al.* 1995). The specificity of a timbre in multidimensional scaling analysis represents its unique, special characteristics that make them dissimilar from the other timbres (Winsberg and Carroll 1989a, 1989b). The specificities are involved in the calculation of the perceptual dissimilarity, or distance, between timbre i and timbre j, d_{ij} as follows (McAdams *et al.* 1995)

$$d_{ij} = \left[\sum_k (x_{ik} - x_{jk})^2 + (s_{ik} + s_{jk})^2 \right]^{\frac{1}{2}}, \tag{8.1}$$

where x_{ik} is the coordinate of the timbre i on the dimension k and s_i is its specificity.

The specificity of the clarinet is always the highest of all natural timbres. This can be noted in Table 8.1, which gives the highest three specificities of natural timbres reported by Krumhansl (1989) and McAdams *et al.* (1995). The high specificity of the clarinet was related to its "unique" spectral feature of the predominance of odd-numbered harmonics (Krumhansl 1989, McAdams *et al.* 1995). I argue that the harp, which has a high specificity, also possesses this feature. This spectral feature, possibly responsible for the characteristic timbre of the harp, is most significant when the string is plucked at the midpoint.

Timbre	Specificity	Timbre	Specificity
Clarinet	25.70	Clarinet	6.4
Harpsichord	22.06	Vibrone	6.4
Harp	18.16	Harpsichord	4.7

Table 8.1: *The highest three specificities of timbres. Krumhansl's (1989) data are shown on the left and McAdams' et al. (1995) on the right row.*

If the high specificities of the clarinet and the harp are both due to the odd-numbered harmonics predominance, this spectral feature is no longer unique to an instrument and the degree of this predominance should be treated as a new dimension. Therefore, the distance between a clarinet tone and a harp tone, d_{12}, should be given by

$$d_{12} = \left[\sum_k (x_{1k} - x_{2k})^2 + (ohp_1 - ohp_2)^2 \right]^{\frac{1}{2}}, \tag{8.2}$$

where *ohp* represents the degree of the odd-numbered harmonics predominance, which can be calculated as, for example, the average amplitude difference between adjacent odd- and even-numbered harmonics in decibel.

Eq. (8.2) is more suited than Eq. (8.1) to estimate the perceptual dissimilarities between the timbres such as the clarinet, the harp, the panpipes, stopped organ pipes, and the *dizi*. Their sound spectra show various degrees of the predominance of odd-numbered harmonics. For these musical instruments, this degree varies through out their tone range. Hence, this dimension is important not only in distinguishing between instruments having jagged or smooth spectral envelopes, but also in differentiating the timbres of the same instrument, as it strongly depends on registers as well as performance techniques, such as the embouchure for the flute (see sections 3.4.2 and 6.6.1) and the plucking point for string instruments.

8.3.2 Grating in harmonic template

Despite the ignorance of spectral envelope jaggedness in psychoacoustics, the auditory system shows a high sensitivity to the microstructure of spectral envelopes. McAdams *et al.* (1999) studied the perceptual importance of various spectrotemporal features by using discrimination paradigms on simplified musical tones. Among the six spectrotemporal simplifications applied to seven instruments tones, they found that the most easily detected was a smoothing of the spectral envelope, accomplished by replacing each harmonic amplitude by the average of itself and its two neighbors.

The pattern of jagged spectra in hollow/nasal sounds has an analogue in visual sensation: the pattern of luminance-varying gratings. This similarity can be demonstrated by transforming a spectrum to a luminance-defined picture. Fig. 8.4 presents two sound spectra of the *dizi* and their corresponding one-dimensional luminance-varying pictures. The luminance of every stripe is determined in terms of a linear function of the harmonic amplitude (in decibel). As this pattern is very distinct in visual world, the odd-numbered harmonics predominance can similarly produce a strong impression, thus playing a role in auditory pattern cognition.

One of the physiological foundations of the high sensitivity to the microstructure of a spectral envelope may be the principle of lateral inhibition in the nervous system. This principle may play a role in detecting a medium predominance of odd-numbered harmonics, such as the *dizi* tone described in Fig. 8.4(b). Traditionally, lateral inhibition in vision is thought to detect and highlight edges and peaks in the special pattern derived from the average firing-rates of the ganglion cells (Ratliff and Hartline 1959). Shamma (2001) suggested that the principle of lateral inhibition might also exist in the auditory nervous system for extracting the shape of spectral envelopes. He proposed the existence of a unified computational framework for central auditory and visual processing. Recent theoretical

views and experiments supported a unified proto-cortical plan for all sensory areas (Dennis and O'Leary 1989, Sur *et al.* 1988). The psychoacoustic finding of a high sensitivity of the auditory system to a spectral envelope smoothing (McAdams *et al.* 1999) also suggested a lateral inhibition network at the level of the harmonic template, which can enhance the amplitude differences between adjacent harmonics, thereby facilitating the detections of spectral envelope jaggedness as well as formants bandwidth. This hypothesis invites further psychophysical and physiological research.

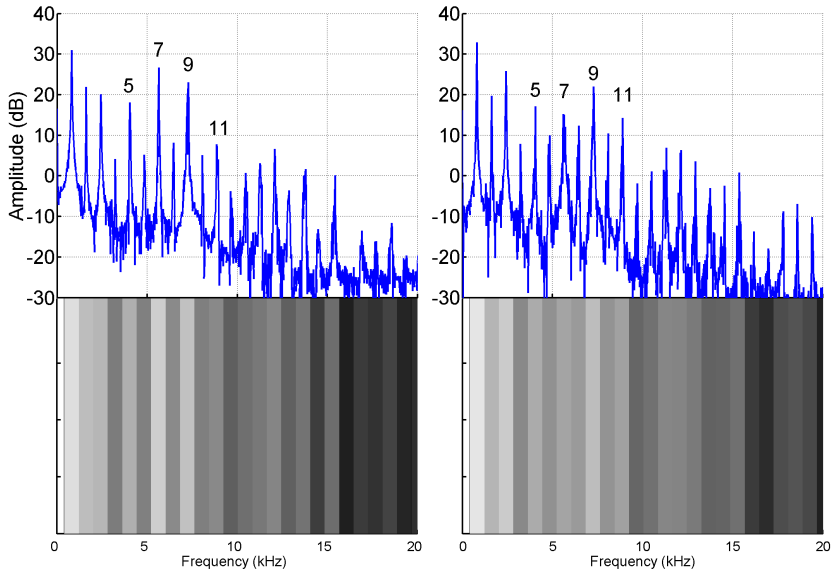


Figure 8.4: *Spectra of two dizi tones (above) and the corresponding luminance-varying pictures (below), where the luminance of each stripe is defined as [(Amplitude in dB)+30]/80. (a) Dizi tone with a high degree of the odd-numbered harmonics predominance. (b) Dizi tone with a medium degree of the odd-numbered harmonics predominance.*

8.3.3 Critical bandwidth correlates of hollowness/nasality

Helmholtz distinguished the auditory attribute correlated to the predominance of low-frequency odd-harmonics from that of high-frequency odd-harmonics. The former elicits the hollow quality and the latter the nasal quality. This differentiation of the two similar sound qualities, reflecting Helmholtz's extraordinary ability to analyze sounds with his ears, can be related to the psychoacoustic aspect of critical bandwidth.

The absence of even-numbered harmonics at high frequencies largely reduces interferences between harmonics in critical bands. At high frequencies, critical bandwidth becomes larger and more harmonics lie in the same critical band. If the even-numbered harmonics at high frequencies are absent, the enlarged spacing between spectral components will reduce unpleasant beatings in critical bands, which always characterize bright musical tones such as trumpet tones. Regarding the rich spectral content inherent in a bright tone, suppressing even-numbered harmonics at high frequencies

may be the most efficient way to reduce interference between harmonics without affecting brightness and pitch. On the other hand, the absence of even-numbered harmonics at low frequencies is unlikely to reduce beatings in critical bands, because critical bandwidths are less than 280 Hz in the frequency range below 2 kHz.

Hollowness and nasality can coexist in a *dizi* tone, as shown in Fig. 8.5. This spectrum compares the resolvability of harmonics in various frequency regions, which is defined as the division of the critical bandwidth and the fundamental frequency (Carlyon and Shackleton 1994), namely the number of harmonics within the same auditory channel. Around 0.5 kHz, the resolvability is 0.2 and the resolvability of odd-numbered harmonics is 0.1. Around 4.5 kHz, the resolvability of harmonics is 1.8 and it decreases to 0.9 when the even-numbered harmonics are absent. The reduction in interference between harmonics is noticeable at high frequencies.

Comparing these *dizi* tone spectra with those of the clarinet, panpipes and the Boehm flute (Fig. 8.1), I suggest that the nasal quality mentioned by Helmholtz is fairly unique to the *dizi*, while the hollow quality is more common in woodwinds and plucked-string instruments. It is interesting to note that the hollow sound quality of low-pitch clarinet tones is sometimes associated with melancholy, whereas the bright *dizi* tones in the second and third registers are described by *dizi* musicians as "clear and sweet", if the odd-numbered harmonics predominate the first formant. The sweet sound quality may be relevant to reductions in interference between high harmonics in critical bands.

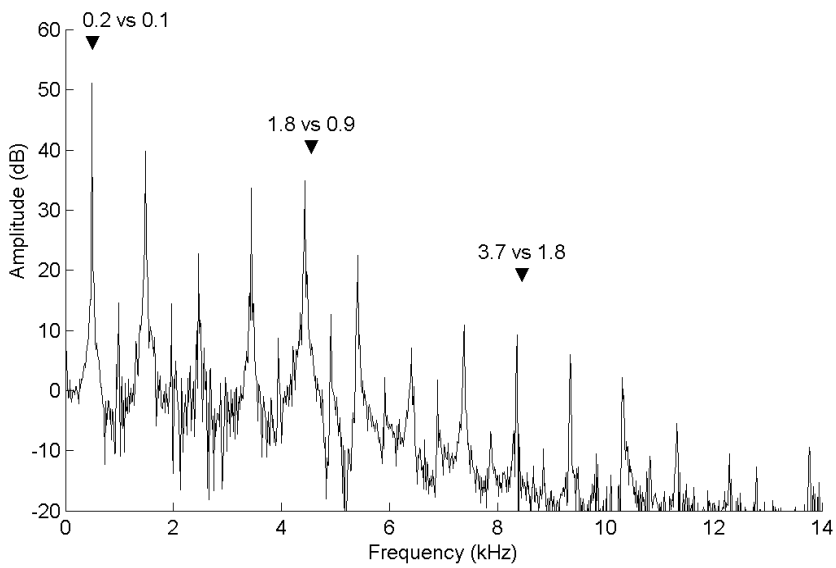
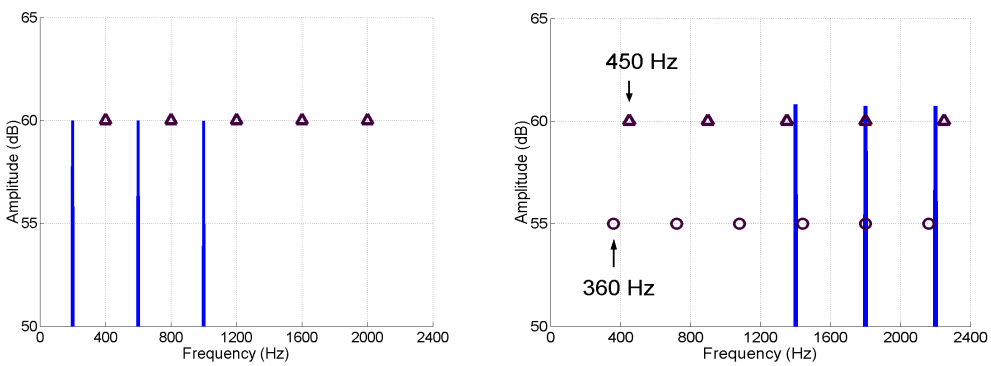


Figure 8.5: *Predominance of odd-numbered harmonics in various frequency regions. Spectrum of the dizi tone B4 (f_0 = 492 Hz) produced by the large dizi with fingering ●○○○○○○, first play mode. The predominance of odd-numbered harmonics in different frequency regions is marked with arrows. The pairs of numbers present the resolvabilities of harmonics for the presence of both even- and odd-numbered harmonics and for the absence of even-numbered harmonics. Resolvability reduction due to the absence of even-numbered harmonics is more significant at higher frequencies.*

8.4 Pitch perception

8.4.1 Residue pitch of double-spaced components

The focus of this section turns from timbre perception to the pitch perception associated with those *dizi* tones having dominant odd-numbered harmonics at low or high frequencies. Benade (1976) reported that the low notes on a clarinet tended to be perceived as an octave higher because the even-numbered harmonics were always fairly weak.[4] It is interesting to examine his observation on sounds with "double-spaced" harmonics in the light of existing models of pitch extraction.



(a) (b)

Figure 8.6: *Residue pitches induced by four consecutive equal-amplitude odd-numbered harmonics of the fundamental frequency f_0 = 200 Hz. (a) Spectrum of the complex tone consisting of 1st, 3rd, and 5th harmonics. The harmonic template ('△') of $2f_0$ = 400 Hz does not fit these three components. (b) Spectrum of the complex tone consisting of 7th, 9th, and 11th harmonics. The harmonic templates of 360 Hz and 450 Hz fit these three components.*

We first consider the "template-based" pitch model relying on spectral cues. This approach assumes that the pitch corresponds to the "harmonic template" that best fits the spectral components of a sound (e.g., Goldstein 1973, Terhardt 1974). Fig. 8.6 shows the spectra of two complex tones consisting of three consecutive odd-numbered harmonics with the fundamental frequency f_0 = 200 Hz. Referring to the "template-based" model of residue pitch, odd-numbered harmonics at low frequencies [Fig. 8.6(a)] cannot induce a pitch $2f_0$ = 400 Hz, because the harmonic template of $2f_0$ does not match the first three odd-numbered harmonics. On the other hand, the complex tone consisting of the 7th, 9th, and 11th harmonics could induce residue pitches of $9f_0/5$ = 360 Hz and $9f_0/4$ = 450 Hz, because their harmonic templates approximately match these upper odd-numbered

[4] Benade appeared to follow Fletcher's (1924) hypothesis that the residue pitch was extracted from distortion products of non-linear cochlear responses such as summation tones and difference tones, which Benade (1976) called "heterodyne components".

harmonics [Fig. 8.6(b)]. Although the upper odd-numbered harmonics are double-spaced, the residue pitches they induced are not exactly equal to $2f_0$ as Benade (1976) suggested. This effect is closely related to the pitch shift of the residue of equally spaced inharmonic components.

The pitch shift of the residue was first noticed by Schouten (1940), but it received little attention until Boer's (1956) dissertation. Boer used complex tones consisting of five inharmonic components spaced 200 Hz apart and found that the pitch shift was a linear function of the frequency shift of these components. A number of experiments were dedicated to this proportionality value (Schouten *et al.* 1962, Patterson 1973, Patterson and Wightman 1976). In Fig. 8.6(b), the pitch shift is 100 Hz and the harmonics frequency shift is 400 Hz, so the proportionality value is 0.25. In Fig. 8.6(c), the pitch shift is 80 Hz and the proportionality value is 0.2. These values replicate Patterson's (1973) observations that the second lowest component dominated in determining the residue pitch extracted from the equal-amplitude inharmonic components when all components are above the fourth harmonic of that pitch. This is also consistent with the "dominance region" of pitch extraction, i.e. that the 3–5 harmonics contribute most to the pitch perception (Ritsma 1967, Plomp 1967).

Although sounds consisting of equally spaced inharmonic components have provided novel stimuli to test models of pitch extraction, such synthesized stimuli have been considered artificial and purely theoretical, as neither musical instruments nor humans can produce a sound composed of equally spaced inharmonic components. The *dizi* tones in the second and third registers bring these studies into practice, because they often have strong 7th, 9th, 11th and 13th harmonics. These "double-spaced" harmonics can evoke a weak pitch sensation besides the fundamental frequency f_0. According to the pitch shift effect, this pitch $2f_0'$ is not exactly an octave above the fundamental frequency; $2f_0' \neq 2f_0$.

8.4.2 Autocorrelation model for pitch extraction and pitch strength estimation

Although the "template-based" pitch model can predict the residue pitch of simple stimuli such as those presented in Fig. 8.6, this purely frequency-domain approach based on pattern matching provides information about the strength of the pitch $2f_0'$ through a complex method of pattern matching and connection weights (Terhardt *et al.* 1982). In general cases I use autocorrelation analysis for pitch extraction and pitch strength estimation.

Temporal correlation functions entail the multiplication of a time-series signal by another signal at different relative time delays. Cross-correlation entails the multiplication of two different signals, while autocorrelation entails the multiplication of a signal by itself. Autocorrelation models of pitch date back to Licklider (1951) and have since been implemented by many others (e.g., Meddis and Hewitt 1991a, Cariani and Delgutte 1996a, Meddis and O'Mard 1997).

Temporal analysis has been encouraged by physiological studies demonstrating *phase locking* of auditory-nerve fiber activity to tone period (see, e.g., Møller 2000). However, a major problem for autocorrelation models of pitch has always been the nature of the central neural processors that analyze temporally-coded information. Such neural elements, whose action would resemble temporal

autocorrelators, have not been observed at any level of the auditory nervous system (Schwarz and Tomlinson 1990), although neural representations of the autocorrelation function were suggested by Cariani and Delgutte (1996a, 1996b). In this sense, the autocorrelation analysis used here should be regarded as a tool of signal processing for demonstrating the phenomenon of multi-pitch in *dizi* tones rather than a physiologically based auditory model.

A standard autocorrelation analysis is used here for pitch extraction. The running autocorrelation of a digital signal *wav*(*t*) at time *t* and lag τ is given by

$$autocorrelation(t, \tau) = \sum_i wav(t + i\Delta t)wav(t + i\Delta t + \tau), \tag{8.3}$$

where Δt is the sample period and the summation is taken for *i* ranging from 1 to $W/\Delta t$ with *W* the width of the time-window. The *pitch-periods* are derived from *peaks* in this autocorrelation function and the *peak heights* reflect the *pitch strengths*.

In order to demonstrate the performance of the pitch model of autocorrelation, three tones with 6–14 harmonics were synthesized. They have different degrees of the predominance of odd-numbered harmonics. Fig. 8.7 shows the spectra and autocorrelation functions of these tones. Their fundamental frequency f_0 = 500 Hz is represented as the main peak at 2 ms in the autocorrelation functions, whereas the two additional pitches $2f_0'$ are represented as the pairs of peaks around 1 ms. It can be observed that the peak heights around 1 ms increase with the amplitude difference between odd- and even-numbered harmonics. This implies that the pitch strengths of the additional pitches $2f_0'$ are sensitive to the average amplitude difference between odd- and even-numbered harmonics. Fig. 8.7 suggests that the additional pitches $2f_0'$ are audible once the average amplitude difference between odd/even-numbered harmonics around the first formant is larger than 8 dB.

Next to the average amplitude difference between odd- and even-numbered harmonics around the first formant, it is also interesting to examine whether the existence of the pitches $2f_0'$ is sensitive to the formant frequency. Fig. 8.8 shows the spectra and the autocorrelation functions of three tones composed of four odd-numbered harmonics accumulated in different frequency ranges. It can be observed that the peak heights around 1 ms increase with the formant frequency. Fig. 8.8 suggests that only the odd-numbered harmonics predominance above 5th harmonics have contribution to the additional pitches $2f_0'$.

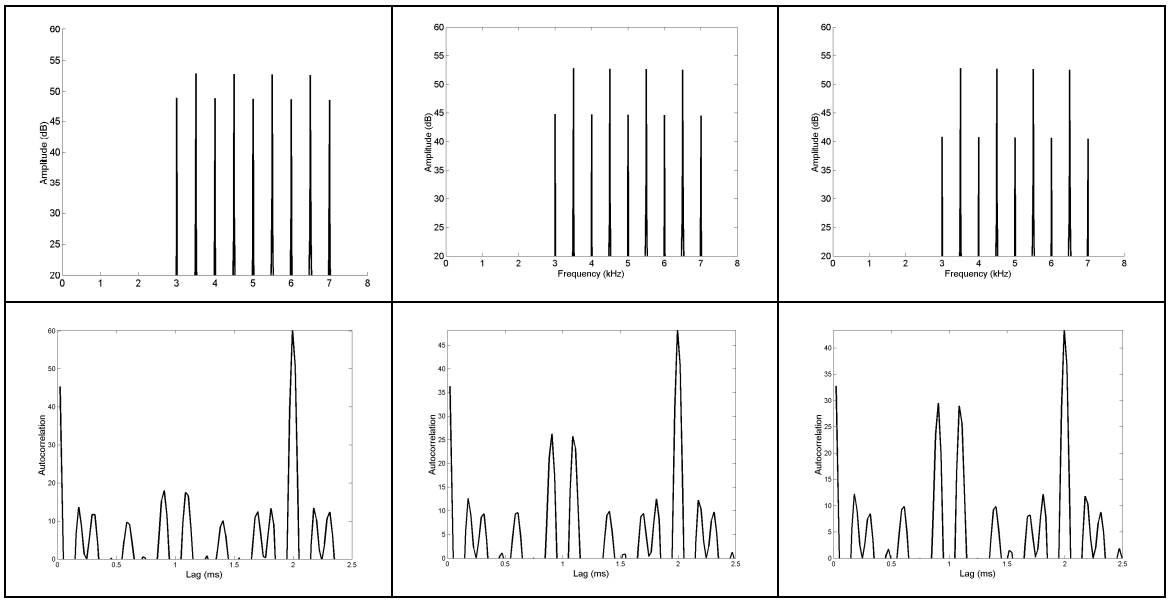


Figure 8.7: *Effect of the amplitude difference between adjacent odd- and even-numbered harmonics on the pitches $2f_0'$. The upper column displays the spectra of three tones with the fundamental frequency f_0 = 500 Hz and with the amplitude difference between odd- and even-numbered harmonics of 4, 8, and 12 dB. The lower column displays their autocorrelation functions.*

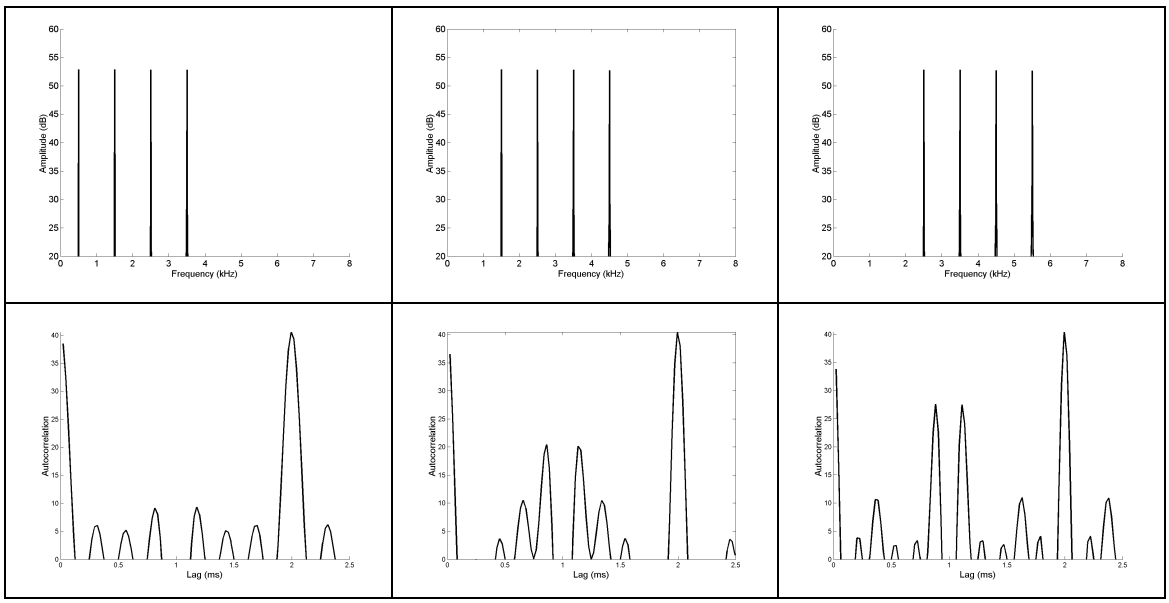
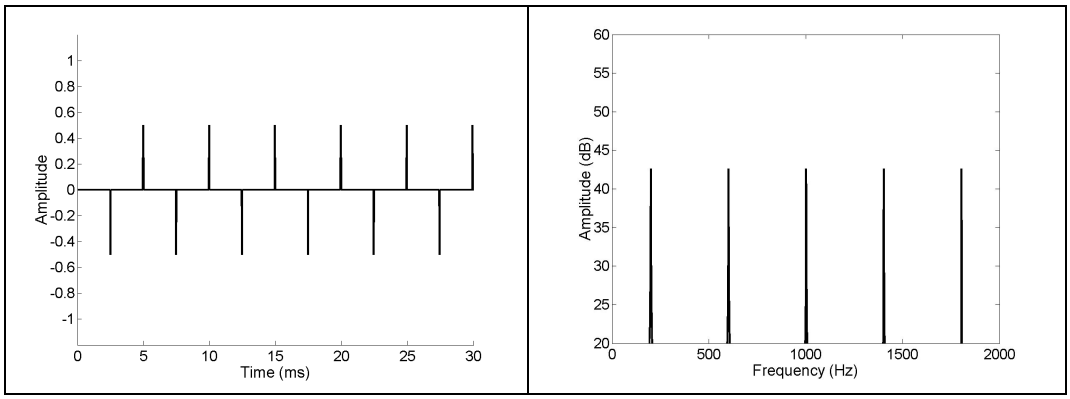


Figure 8.8: *Effect of the center frequency of the formant of odd-numbered harmonics on the pitches $2f_0'$. The upper column displays the spectra of three tones composed of four consecutive odd-numbered harmonics (f_0 = 500 Hz). The lower column displays their autocorrelation functions. The peak heights around 1 ms increase and their distances decrease as the formant frequency increases.*



(a) (b)

Figure 8.9: *Alternative-polarity clicks. (a) Waveform. (b) Spectrum.*

The present study on the pitch perception associated with the odd-numbered harmonics predominance is closely related to the phenomenon of rate pitches of alternative-polarity clicks, whose spectra show infinite equal-amplitude odd-numbered harmonics without even-numbered harmonics (Fig. 8.9). When these click trains are high-pass filtered, rate pitches are strengthened (Flanagan and Gutman 1964). Cariani and Delgutte (1996b) explained the existence of rate pitches of alternative-polarity clicks in terms of their autocorrelation functions, which show split peaks around the rate interval. Based on the same argument, the multi-pitches of *dizi* tones with a predominance of odd-numbered harmonics also represent as split peaks in their autocorrelation functions. However, the pitch perception in *dizi* music is much more complex than that of filtered clicks, because the degree and the frequency region of the odd-numbered harmonics predominance varies a lot in *dizi* music and it is not straightforward to determine which pitches are audible. The phenomenon of multi-pitch makes the transcription of *dizi* solo music fairly challenging, as will be demonstrated in the following example.

8.4.3 Multi-pitches in *dizi* solo music

Although dominant 7th, 9th, and 11th harmonics of a *dizi* tone are predicted to produce pitches $2f_0'$, their pitch strengths seem fairly low. I myself can hear "something" above the fundamental frequency, but cannot judge their exact musical pitches. To determine the multi-pitches of *dizi* tones, an autocorrelation analysis for transcription is used. A transcription example is given in Figs. 8.10 and 8.11, where notes are determined according to some "streams" in the autocorrelogram, which has its *y*-axis in the cent scale to match the lines of the soprano clef. It is important to note that these plots, representing the values of the running autocorrelation function, not necessarily corresponds to the perceived pitch strength, because there are repetitive patterns in an autocorrelation function. Details of the generation method of this autocorrelogram are described in Appendix A.

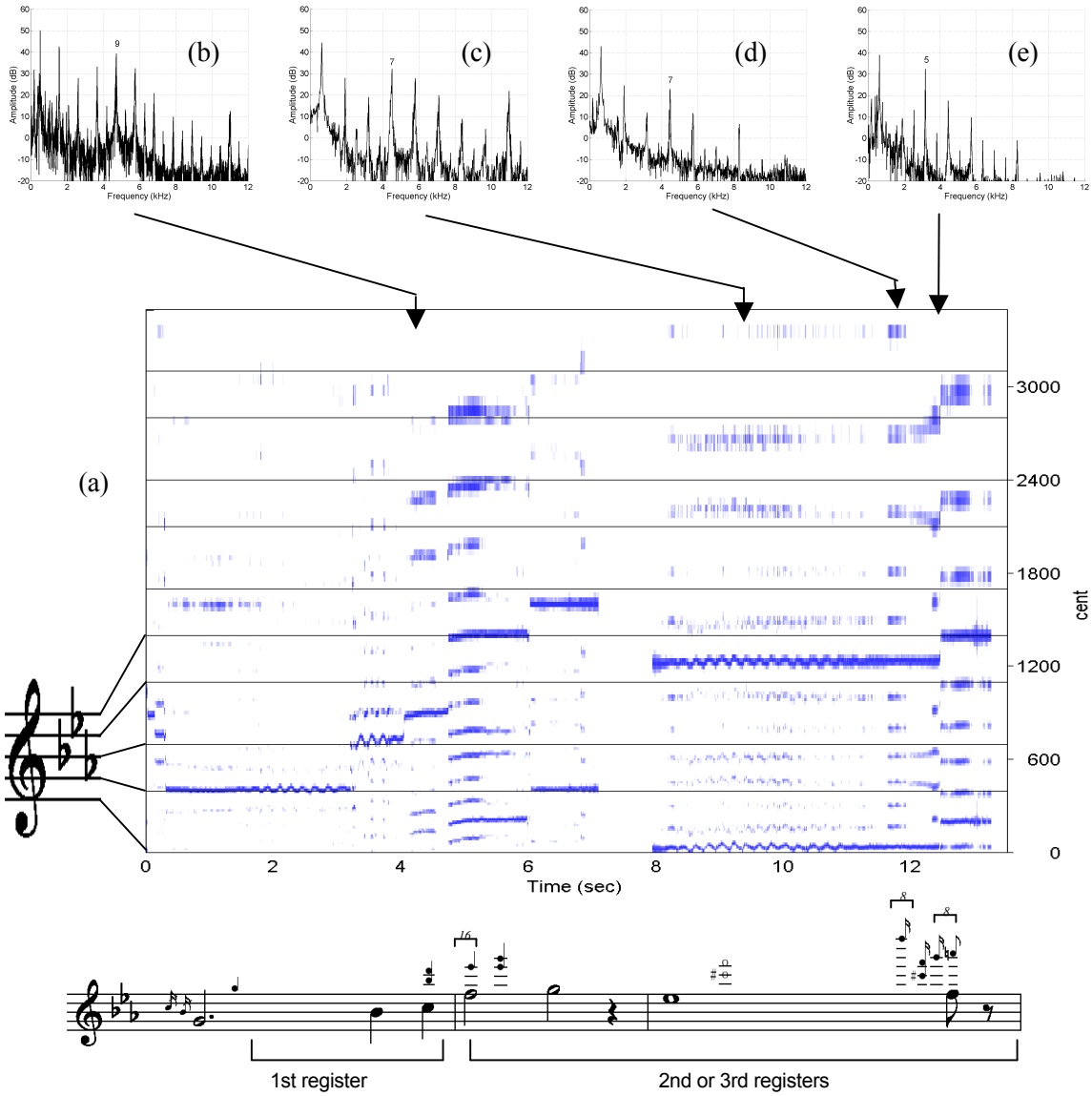


Figure 8.10: *Multi-pitches in dizi solo music (excerpt from "Yiolan Fongchun") represented in autocorrelogram. (a) Autocorrelogram. (b) Spectrum snapshot at 4.2 sec. The dominant 7th, 9th, and 11th harmonics induce the pitches B♭6 and D7. (c) Spectrum snapshot at 9 sec. (d) Spectrum snapshot at 11.7 sec. (d) Spectrum snapshot at 12.3 sec.*

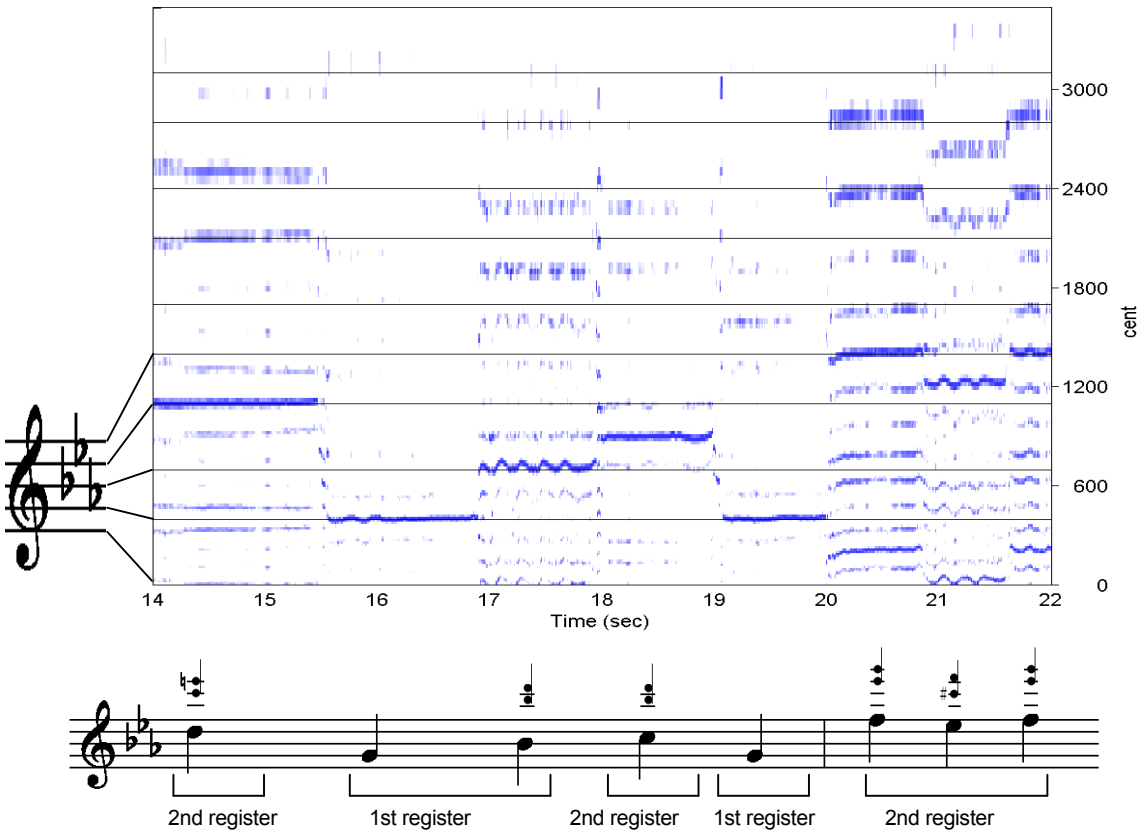


Figure 8.11: *Multi-pitches in dizi solo music (continued).*

Fig. 8.10 shows three types of multi-pitches. First, the autocorrelogram suggests that at 1 sec there is a weak pitch G6 above the prominent pitch G5. This corresponds to the spectral feature of a predominance of the even-numbered harmonics (for the spectrum, see Fig. 6.18). The melody ascends from the first register to the second register at 4 sec with the appearance of the second type of multi-pitches: $9f_0/4$ and $9f_0/5$ induced by dominant 7th, 9th and 11th harmonics [Fig. 8.10(b), see also Fig. A.4]. At 5 sec, the prominent pitch F6 is accompanied by the pitch of its strong 7th harmonic, which is segregated from F6 by inducing a pitch of $7f_0$. This belongs to the third type of multi-pitches: nf_0.

The long *dizi* tone E6 has a significant characteristic of multi-pitches $2f_0'$, as the spectrum show dominant 7th, 9th and 11th harmonics [Fig. 8.10(c)]. At 11.7 second, the 7th harmonics becomes so outstanding that its pitch $7f_0$ is likely to replaces the multi-pitches $2f_0'$ [Fig. 8.10(d)]. At the end of this tone, the 5th harmonic becomes outstanding so that the pitch $5f_0$ could be audible [Fig. 8.10(e)].

From 14 to 22 second, all *dizi* tones in the second register have the multi-pitches of $2f_0'$, which are represented in the transcription as major thirds, moving parallel to the prominent pitch f_0 (Fig. 8.11). However, it seems difficult to judge their exact musical pitches by listening.

Although the method of autocorrelogram provides a valuable tool for estimating the pitches and

their strengths in *dizi* solo music, there is a gap between the calculated multi-pitches and perceived multi-pitches. From a perspective of psychoacoustics, the autocorrelogram representations of multi-pitches demand intensive experimental studies because of low pitch strengths. Moreover, the method of autocorrelogram for multi-pitch transcription of *dizi* solo music lacks a physiological basis for auditory temporal coding at high frequencies. The present autocorrelation representation extends up to the limits of phase locking in the auditory nervous system. Therefore, a new model fusing the temporal and the place representation for extracting pitches from high-frequency (> 4 kHz) harmonics is needed. In spite of debates over the respective roles of temporal and place representations in the perception of pitch, it should be noted that the multi-pitches of $2f_0'$ are also predicted by a place theory (see Fig. A.4).

The multi-pitch sensation evoked by a predominance of high odd-numbered harmonics sheds new lights on the aspects of auditory scene analysis and timbre perception. It is generally accepted that a quasi-periodic complex tone consisting of frequencies nf_0 evokes one single pitch f_0. Yet, *dizi* tones with dominant odd-numbered harmonics around the first formant (4–6 kHz) can evoke two additional pitches $2f_0'$, which are unlikely to blend with the prominent pitch f_0 because $2f_0' \neq 2f_0$. Although their exact musical pitches seem perceptually ambiguous, they may play a role in timbre perception through stream segregation. I suggest that *dizi* tones with dominant odd-numbered harmonics around the first formant can evoke multi-streams in the auditory scene. This phenomenon is comparable to the multi-streams in tones of overtone-singing and Jew's harp, which are characterized by a strong harmonic segregated from the fundamental frequency. However, the multi-pitch phenomenon of *dizi* tones with dominant odd-numbered harmonics at high frequencies is rather special, because there is no spectral component at $2f_0'$, and maximally one component exactly at $2nf_0'$.

8.5 Conclusions

Dizi tones in the second and third registers, characterized by weak even-numbered harmonics, have the hollow and/or nasal sound quality described by Helmholtz. The distinct timbre evoked by the spectral feature of large amplitude differences between odd-even-numbered harmonics is explained by analogizing it as the visual pattern of a luminance-contrasting grating in a harmonic template.

The odd-numbered harmonics dominating at low and high frequencies have different perceptual effects. When this predominance occurs at high frequencies, interference between adjacent harmonics is reduced. This may explain that nasal *dizi* tones are described as sweet by musicians. Moreover, the predominance of odd-numbered harmonics at high frequencies can induce additional pitches $2f_0'$ that enrich the texture in *dizi* solo music. An autocorrelation model of pitch suggests two conditions for the existence of the additional pitches: (1) odd-numbered harmonics must dominate in the frequency range above the 5th harmonics, and (2) the average amplitude difference between odd/even-numbered harmonics in this range must be larger than 8 dB. As these two conditions are

often satisfied in the *dizi* tones in the second and third registers but not in most tones produced by other musical instruments, this multi-pitch phenomenon is fairly unique to the *dizi*.

Autocorrelation analysis of *dizi* tones with dominant odd-numbered harmonics suggests that the additional pitches $2f_0'$ are major thirds moving parallel to the prominent pitch of the fundamental frequency f_0. This prediction invites further psychoacoustic studies.

References

Adler, S. (2002). The Study of Orchestration. W.W. Norton and Company.

Benade, A.H. (1976). Fundamentals of Musical Acoustics. Oxford University Press.

Cariani, P.A., and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. Journal of Neuro-physiology 76, 1698–1716.

Cariani, P.A., and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. Journal of Neurophysiology 76, 1717–1734.

Carlyon, R.P., and Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? Journal of the Acoustical Society of America 95, 3541-3554.

Chen, M.Y. (1997). Acoustic correlates of English and French nasalized vowels. Journal of the Acoustical Society of America 102(4), 2360-2370.

Dang, J., and Honda, K. (1996). Acoustic properties of paranasal sinuses derived from transmission characteristic measurement and morphological observation. Journal of the Acoustical Society of America 100, 3374–3383.

de Boer, E. (1956). On the "Residue" in Hearing. Unpublished PhD thesis, University of Amsterdam.

de Boer, E. (1976). On the "residue" and auditory pitch perception. In: Handbook of Sensory Physiology (edited by W.D. Keidel and W.D. Neff), Vol.3, Springer-Verlag, 479-583.

de Cheveigné, A. (1998). Cancellation model of pitch perception. Journal of the Acoustical Society of America 103, 1261-1271.

Dennis, D., and O'Leary, M. (1989). Do cortical areas emerge from a protocortex? Trends in Neurosciences 12,400-406.

Fant, G. (1960). The Acoustic Theory of Speech Production. Mouton, The Hague.

Flanagan, J.L., and Gutman, N. (1964). Pitch of high-pass-filtered pulse trains. Journal of the Acoustical Society of America 36, 757-765.

Fletcher, H. (1924). The physics criterion for determining the pitch of musical tone. Physics Review 23, 427-437.

Fletcher, N.H., and Rossing, T.D. (1991). The Physics of Musical Instruments. Springer-Verlag.

Gibson, O.L. (1994). Clarinet Acoustics. Indiana University Press.

Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex

tones. Journal of the Acoustical Society of America 54, 1496-1516.

Heike, G. (1982). Phonologie. Metzler, Stuttgart.

Helmholtz, H.L.F. (1877). On the Sensations of Tone, 4th ed., translated by A.J. Ellis, Dover, New York, 1954.

House, A.S., and Stevens, K.N. (1956). Analog studies of the nasalization of vowels. Journal of Speech and Hearing Dis. 21, 218–232.

Iverson, P., and Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America 94, 2595–2603.

Jost, E. (1967). Akustische und psychometrische Untersuchungen an Klarinettenklängen (= Veröffentl. d. Staatl. Inst. f. Musikforschung Berlin, Band 1), Köln (Arno Volk) 1967.

Krimphoff, J., McAdams, S., and Winsberg, S. (1994) Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychophysique. Journal de Physique IV, Colloque C5, Vol. 4.

Krumhansl, C.L. (1989). Why is musical timbre so hard to understand? In: Structure and Perception of Electroacoustic Sound and Music (edited by S. Nielzen and O. Olsson), Excerpta Medica, 43-53.

Licklider, J.C.R. (1951). A duplex theory of pitch perception. Experientia 7(4), 128-134.

McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., and Krimphoff, J. (1992). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research 58, 177-192.

McAdams, S. (1993) Recognition of Auditory Sound Sources and Events. In: Thinking in Sound, Oxford: Oxford University Press.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes, Psychological Research 58, 177-192.

McAdams, S., Beauchamp, J.M., and Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. Journal of the Acoustical Society of America 105(2), 882–897.

Meddis, R., Hewitt, and M.J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. Journal of the Acoustical Society of America 89, 2866–2882.

Meddis,R., Hewitt, and M.J. (1991b).Virtual pitch and phase sensitivity of a computer model of the auditory periphery.II. Phase sensitivity. Journal of the Acoustical Society of America 89, 2883-2894.

Meddis, R., and O'Mard L. (1997). A unitary model of pitch perception. Journal of the Acoustical Society of America 102, 1811-1820.

Møller, A.R. (2000). Hearing: its Physiology and Pathophysiology. Academic Press.

Patterson, R.D. (1973). The effects of relative phase and the number of components on residue pitch.

Journal of the Acoustical Society of America 53, 1565-1572.

Patterson, R.D., and Wightman, F.L. (1976). Residue pitch as a function of component spacing. Journal of the Acoustical Society of America 59, 1450-1459.

Plomp, R. (1967). Pitch of complex tones. Journal of the Acoustical Society of America 41, 1526-1533.

Preis, A. (1984). An attempt to describe the parameter determining the timbre of steady-state harmonic complex tones, Acustica 55, 1-13.

Ratliff, F., and Hartline, H.K. (1959). The responses of limulus optic nerve fibers to patterns of illumination on the receptor mosaic. Journal of General Physiology 42, 1241-1255.

Ritsma, R.J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. Journal of the Acoustical Society of America 42, 191.

Rusko, M. (1997). Toward a more general understanding of the nasality phenomenon. In Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology, (Ed. M. Leman, Berlin: Springer) 351-361.

Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962). Pitch of the residue. Journal of the Acoustical Society of America 34, 1418-1424.

Schwarz, D.W.F., and Tomlinson, R.W.W. (1990). Spectral response patterns of auditory cortical neurons to harmonic complex tones in alert monkey (Macaca mulatta). Journal of Neurophysiology 64, 282-298.

Shamma, S. (2001). On the role of space and time in auditory processing. Trends in Cognitive Science 5(8), 340-348.

Sur, M., Garraghty, P.E., and Roe, A.W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. Science 242,1437-1441.

Terhardt, E. (1974). Pitch, consonance, and harmony. Journal of the Acoustical Society of America 55, 1061-1069.

Terhardt, E., Stoll G., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. Journal of the Acoustical Society of America 71(3), 679-688.

Wessel, D. (1979) Timbre space as musical control structure. Computer Music Journal 3, 45-52.

White, L.J., and Plack, C.J. (1998). Temporal processing of the pitch of complex tones. Journal of the Acoustical Society of America 103 (4), 2051-2059.

Winsberg, S., and Carroll, J.D. (1989a). A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSCAL model. In: Multiway data analysis (edited by R. Coppi and S. Belasco), North Holland-Elsevier, 405-14.

Winsberg, S., and Carroll, J.D. (1989b). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. Psychometrika 54, 217-29.

Zhao, S.T. (1985). *Diyi Chunqio*.（趙松庭《笛藝春秋》，浙江人民出版社）

Chapter 9.
Perception of subharmonics: Roughness

Abstract

Subharmonics at $(2n\text{-}1)f_0/2$ (where f_0 is the perceived pitch, n = 1, 2, 3...) sometimes appear in the *dizi* tones in the second octave. They do not produce another pitch one octave below, but merely induce a rough sound quality. Subharmonics are common in human vocalization and have been identified as a source of roughness. However, roughness induced by subharmonics cannot be explained by past psychoacoustic models and appears against the theory of consonance-dissonance. Roughness induced by subharmonics can be explained by taking into account the pitch-based grouping mechanism that integrates information across auditory channels. A qualitative model is proposed with two assumptions: (1) grouping components at nf_0 demands a robust pitch sensation of f_0; (2) unpleasant beatings caused by components in the same critical bands are largely reduced by this grouping. Since subharmonics in *dizi* tones at low frequencies are always much weaker than adjacent harmonics, the subharmonic pitch of $f_0/2$ cannot be evoked. Therefore, the ungrouped subharmonics are perceived as impurities of *dizi* tones.

A large *dizi* with a slack membrane has a strong tendency to produce rich subharmonics. Because the tone range of a large *dizi* is comparable to human singing, it is especially optimal for imitating human voices. Large *dizi* tones with subharmonics often recall melancholic voices.

9.1 Introduction

Chapters 7 and 8 discussed the perception of harmonics of *dizi* tones. This chapter accounts for the perception of their subharmonics, which appear between spectral lines of harmonics in the spectrum. The frequencies of subharmonics are multiples of a fraction of the fundamental frequency f_0. Subharmonics with frequencies $(2n\text{-}1)f_0/2$ can appear in *dizi* tones in the second play mode, while subharmonics with frequencies $(3n\text{-}1)f_0/3$ and $(3n\text{-}2)f_0/3$ can appear in *dizi* tones in the third play mode.

The occurrence of subharmonics is of perceptual importance only for membrane flutes, because membrane-less flutes are not able to produce high-frequency subharmonics (see Fig. 6.5). For the *dizi* tones played in the third mode, the membrane is always located near a pressure node of the fundamental, so that the *dizi* behaves like a membrane-less flute. Subharmonics with frequencies $(3n\text{-}1)f_0/3$ and $(3n\text{-}2)f_0/3$ are thus very weak at high frequencies and have little perceptual effect. For this reason, this chapter focuses on the perception of the subharmonics with frequencies $(2n\text{-}1)f_0/2$ in the *dizi* tones in the third register.

In comparison with harmonics, subharmonics in musical tones have attracted little attention from musical acousticians. While physicists and musicians have explored subharmonics produced by bowed string instruments (Lauterborn 1996, Kimura 1999), brass and reed instruments (Gibiat and Castellengo 2000), no research on timbre spaces has mentioned the contribution of subharmonics to musical timbres.

The perception of subharmonics in natural sounds has mainly been studied within the framework of speech research. Voice clinicians have identified subharmonics as a source of roughness in human voices (e.g., Omori *et al.* 1997). In the field of music perception, stationary violin tones with subharmonics were related to an auditory attribute similar to roughness: rustle (Stepanek and Otcenasek 1999). However, few investigations have related the perception of rough voices with subharmonics to psychoacoustic models of roughness. A computational study based on Aures' (1985) model of roughness showed a medium correlation between computed roughness and perceived roughness. This result highlighted the gap between psychoacoustic research using synthesized stimuli and the perception of natural rough sounds such as pathological human voices.

The goal of this chapter is to indicate the failure of past psychoacoustic models and to provide a new model based on a grouping mechanism in auditory scene analysis. Section 9.2 gives a general review of auditory roughness and difficulties of relevant psychoacoustic models. Section 9.3 addresses the methods and results of a preliminary experiment, which was designed to show that no existing psychoacoustic models could explain roughness induced by subharmonics. A qualitative model of roughness induced by subharmonics is described in section 9.4. Section 9.5 discusses the role of subharmonics in *dizi* music.

9.2 Roughness: models and difficulties

9.2.1 Psychoacoustic models and neurophysiological evidence

In psychoacoustics, roughness is an important parameter that induces unpleasant qualities of a sound. Since its introduction by Helmholtz (1877), roughness has been considered to be due to rapid beatings in auditory channels, or critical bands. He explained the perception of dissonances in Western music in terms of roughness evoked by adjacent harmonics of two simultaneously sounding musical tones. The consonant relations were viewed as exceptional cases, when either no or very few

beatings were produced.

Psychoacoustic experiments on roughness have been performed using various types of synthesized sounds. Using simultaneous pure-tone pairs stimuli, Plomp and Levelt (1965) plotted an idealized curve of the relationship between consonance and critical bandwidth. This curve serves as the basis of curve-mapping models of roughness, which are confined in the frequency-domain.

In contrast to curve-mapping models, auditory models of roughness rely on auditory processing and provide an explanatory model for roughness. Several calculation procedures have been developed (Aures 1985, Daniel and Weber 1997, Leman 2000), which work in both the frequency- and time-domains. The former corresponds to the activation patterns on the basilar membrane, while the latter corresponds to the total firing rate of auditory nerves, which is phase-locked to stimulus components in auditory channels (see, e.g., Møller 2000).

Recent progress on the search for the physiological basis of roughness mainly concerned temporal coding at various levels of the auditory nervous system. Tramo *et al.* (1992, 2000) found a correlate of roughness in the temporal discharge patterns of auditory nerve fibers. McKinney *et al.* (2001) explored the response of neurons in inferior colliculus to musical intervals. They found that dissonance correlated qualitatively with rate fluctuations of inferior colliculus neurons, demonstrating Helmholtz's (1877) hypothesis that dissonance tended to induce intermittent responses in auditory nerve fibers.

Neural representation of roughness has also been found in higher levels of the auditory pathway, including the auditory cortex (for a review, see Langner 1992). Physiological studies of responses to periodic stimuli in the primary auditory cortex have generally reported or implied limiting rates of phase-locking that are considerably lower than psychoacoustic values of the upper limit of roughness perception (e.g., Gaese and Ostwald 1995, Eggermont 1998). Suggesting that this discrepancy may stem from the use of anesthetized animals, Fishman *et al.* (2000) measured neural responses in primary auditory cortex of an *awake* macaque monkey. They used stimuli consisting of three consecutive harmonics of fundamental frequency f_0 ranging from 25 to 4000 Hz and found that in the thalamo-recipient zone, the stimulus f_0 at which phase-locking was maximal increased with the best frequency and reached an upper limit of between 75 and 150 Hz for the best frequency greater than about 3 kHz. These physiological relationships parallel psychoacoustic data that showed the roughness of an amplitude-modulated tone with a carrier frequency 1 kHz reached its maximum near modulation frequencies of 70 Hz (Zwicker and Fastl 1990). However, such phase-locked firing patterns should be explained no more than maintenance of temporal information at lower levels of auditory processing. Its correlation to the unpleasant sensation associated with dissonance or roughness, which is likely coded in the prefrontal cortex, remains to be demonstrated.

Tramo *et al.* (2001) studied the effects of auditory cortex lesions on the discrimination of roughness, consonance-dissonance and pitch. A patient with infarcts in the auditory cortex, MHS, had difficulty in consonance-dissonance discrimination, but his performance of roughness discrimination of two simultaneous pure tones was normal. As MHS also had difficulty in frequency discrimination,

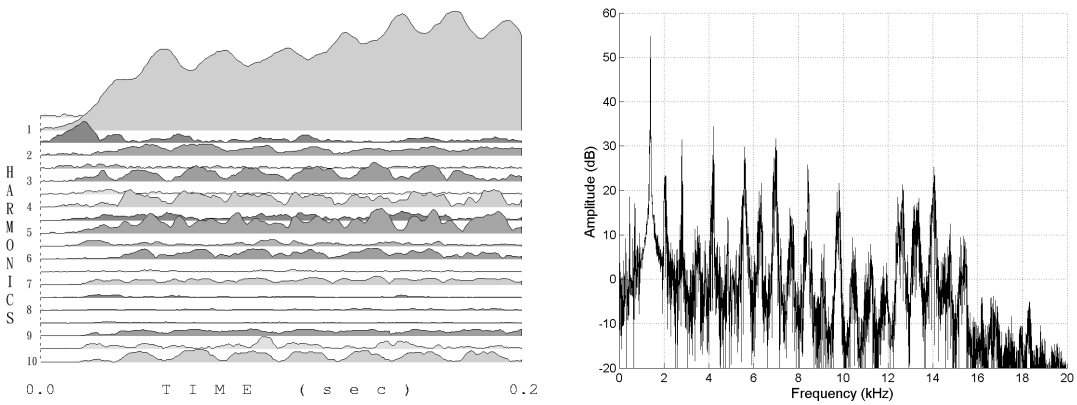
it was concluded that MHS's impaired consonance-dissonance perception was due to impairment in pitch perception but not roughness perception. The authors thus suggested that pitch relationships influence harmony perception in the vertical dimension with roughness playing a secondary role.

In summary, some of recent physiological data recorded at various levels of the auditory pathway have been interpreted as neural representations of psychoacoustic roughness. However, the correlation between the fluctuating, phase-locked neural activity and the unpleasant sensation of roughness remains to be clarified. A case study of brain lesion in the auditory cortex showed a disassociation between harmony perception and roughness, thus invalidating both Helmholtz's theory of dissonance and neural representations of roughness.

9.2.2 Pleasant beatings in *dizi* music

As the aspect of roughness was developed to explain the perception of harmony in Western music, it would not be surprising that a comparative approach can give evidence against a universal, biologically-based explanation of the sensation of roughness. The perception of amplitude-modulated sounds can be culture- and context-dependent, as pleasant beating-like sounds prevail in non-Western music. For instance, beatings in gong tones are essential in gamelan music (Gomperts 1995).

In flute music, beating-like sounds can easily be produced with flutter tonguing. Fig. 9.1(a) shows the amplitude envelopes of harmonics and subharmonics of a flutter tongued *bangdi* tone. The amplitude envelope of the fundamental shows 25% amplitude modulation with the modulation frequency of 40 Hz. Most of other components have the same modulation frequency but are nearly 100% modulated. This effect, in line with Worman's theorem (section 5.3), is also represented as the widening of spectral lines of high-frequency components in Fig. 9.1(b).



(a) (b)

Figure 9.1: *Amplitude-modulations in a fluttering tongued flute tone produced by the bangdi in F. The fundamental frequency is about 1397 Hz. (a) Amplitude envelop of the lowest 10 harmonics and subharmonics of this tone. The amplitude modulation frequency of the fundamental is about 40 Hz. (b) Spectrum of this tone.*

Although the physical mechanism correlated to different amounts of amplitude-modulation of harmonics is unknown, from a psychoacoustic perspective, it is clear that a high value of auditory roughness is inherent of fluttering tongued *dizi* tones. The fact that amplitude-modulated *bangdi* tones sound fairly pleasant and exciting for Chinese audience places current psychoacoustic theories of roughness in doubt. It is very important to note that in Western flute music, the use of flutter tonguing tends to be confined to contemporary music, whereas it is one of the backbone techniques in the Northern style of *dizi* music. As fluttering tongued flute tones are more favored in membrane flute music, amplitude-modulated high harmonics produced by the membrane may play a role in evoking a pleasant sensation in a beating-like tone.

In traditional *dizi* music the use of flutter tonguing is restricted in the *bangdi* (Chinese "piccolo"), while musicians consider flutter tongued tones of larger *dizi* as "ineffective". It seems that the spacing between harmonics – the pitch – is important in the perception of fluttering tongued flute tones. This might be related to that widened spectral lines of harmonics start to overlap each other when the fundamental is lower than a threshold (f_0 < 400 Hz, for instance). A "dirty" timbre associated with low-pitch tongued *dizi* tones may arise from a spectral "continuum" at high frequencies.

9.2.3 Pleasant beatings in voices

In addition to amplitude-modulated musical tones, another example of "pleasant beats" in auditory channels can be found in low-pitch male voices. According to psychoacousticians, a voice with fundamental frequency below 200 Hz can elicit roughness, since its adjacent upper harmonics can induce fluctuation of the temporal envelop in auditory channels. However, such rich-harmonics low-pitch voices are not rough according to voice clinicians and audiences. In singing practice, the voices of a professional bass singer are characterized by a strong formant at about 3 kHz, the singer's formant. In this frequency region, the critical bandwidth is about 400–600 Hz. Consequently, if the pitch of a singing voice is below 200 Hz, the harmonics unresolved by auditory channels could give the voice a rough sound quality. Regarding this unpleasant quality, a bass should reduce the strength of the singer's formant when he sings low notes. However, no such a technique has been reported. By contrast, the audience favors bright bass voices rather than rather than dull ones that are free of roughness.

9.2.4 Unpleasant consonance: Roughness induced by subharmonics

Despite a plethora of calculation models of roughness, these models encounter difficulties when applied to human voices. Roughness as an indicator of pathological voices has been extensively studied and scored by clinicians. A serious conflict between psychoacoustic and clinical studies of human voices can be found in a specific type of roughness: roughness induced by subharmonics.

Since the discovery of subharmonics in human voices (Dejonckere and Lebacq 1983), this phenomenon has been related to the period doubling scenario in non-linear dynamical systems (Feigenbaum 1978) and has been studied through two-mass models of the glottis (Herzel 1993,

Steinecke and Herzel 1995, Mergell *et al.* 2000).

Subharmonics in human voices shed new light on the perception of pitch and roughness. When subharmonics with frequencies $(2n\text{-}1)f_0/2$ appear in a voice, they seldom make the pitch an octave lower. Instead, they merely introduce a rough sound quality to it. Fig. 9.2 gives the waveform of a human voice around the period-doubling bifurcation, comparing the spectra before and after the bifurcation. Perceptually, the period-doubling bifurcation introduced a rough quality to this voice. It should be noted that low subharmonics are noticeably weaker than adjacent harmonics, whereas high subharmonics are comparable to their adjacent harmonics in magnitude. This spectral characteristic also exists in *dizi* tones with subharmonics [see Figs. 3.6(b), 6.5 (b), 6.13(c), 9.1(b), 9.8(a), 9.9].

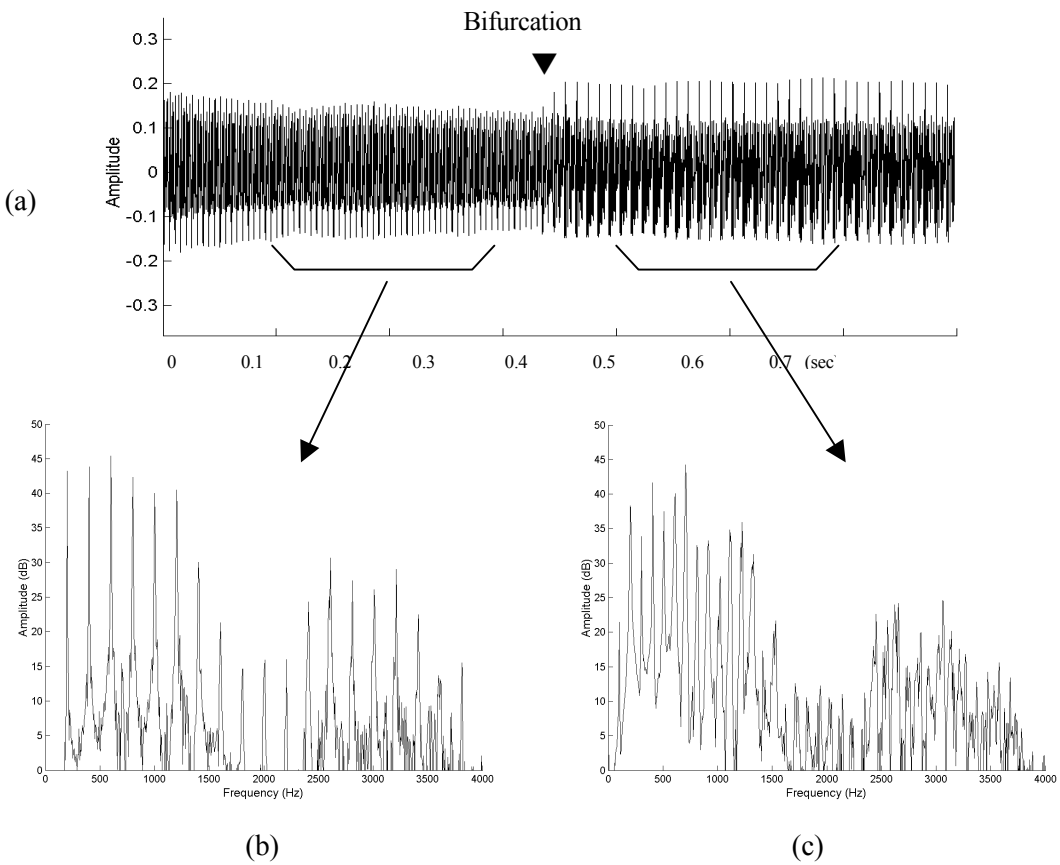


Figure 9.2: *Subharmonics in a human voice. (a) Waveform of this voice around a period-doubling bifurcation. (b) Spectrum of the voice before the bifurcation. (c) Spectrum of the voice after the bifurcation. Subharmonics emerge at midpoints of pairs of adjacent harmonics. Note that the lowest two subharmonics are noticeably weaker than adjacent harmonics.*

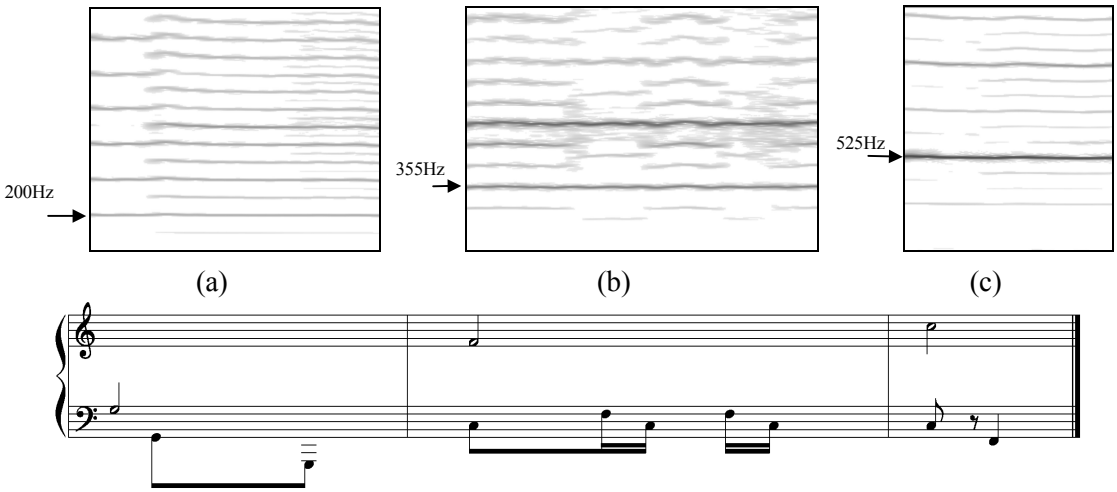


Figure 9.3: *Human voices with subharmonics. (a) G3. (b) F4. (c) C5.*

Fig. 9.3 gives three spectrograms of human voices with subharmonics and the corresponding musical notes. The prominent pitch f_0 is represented as the lowest strong spectral component in each spectrogram (marked with an arrow). The space between harmonics (with frequencies nf_0) is divided by subharmonics. In Fig. 9.3(a), the frequencies of all subharmonics $(2n\text{-}1)f_0/2$ are integer multiples of the "subharmonic fundamental" $f_0/2$. Therefore, a sudden appearance of these subharmonics could be regarded as adding a tone one octave below to the voice. Similarly, the appearance of frequencies $(3n\text{-}1)f_0/3$ and $(3n\text{-}2)f_0/3$ could be regarded as the addition of a tone a twelfth below [Fig. 9.3(b)].

Roughness induced by subharmonics seriously contradicts the music theory of consonance-dissonance. The frequencies of subharmonics in a voice are always multiples of a low integer fraction of the fundamental frequency, such as $f_0/2$, $f_0/3$, $f_0/4$, even $f_0/6$ [Fig. 9.3(c)]. As the corresponding subharmonic pitches and the fundamental pitches strictly stand in simple integer ratios $1{:}n$, it is a puzzle that such "consonances" in voices are characterized by a rough sound quality.

9.2.5 Failure of psychoacoustic models in roughness calculation

In phoniatrics, several indicators have been used for describing pathological voice quality, including roughness, breathiness and hoarseness. It is interesting to compare psychoacoustic roughness with the evaluation of roughness in speech research. Roughness has been related to 'aperiodicity features' of human voices (e.g., Arends *et al.* 1990, Dejonckere 1995), which arise from the glottis' instability. Changes in the waveform shape between glottal cycles can be measured by the correlation coefficient calculated for all pairs of successive cycles. Various perturbation measures have been provided, such as jitter (variations of pitch) and shimmer (variations of amplitude). This perturbation measures approach parallels the evaluations of psychoacoustic roughness in amplitude-modulated tones and frequency-modulated tones.

Roughness evaluations of human voices have provided new data for examining psychoacoustic models of roughness. In Reuter's (2000) dissertation, both perturbation analysis and a psychoacoustic

model of roughness were applied to pathological voices which show subharmonics, biphonation or deterministic chaos. The degrees of roughness, breathiness and hoarseness of these voices had been rated by clinicians. Computer programs based on perturbation analysis and a psychoacoustic model (Aures 1985) of roughness were written to compute roughness of these voices. It was a striking finding that both the perturbation analysis and the psychoacoustic model performed poorly; computed results showed a medium correlation to the perceived roughness.

The empirical finding of roughness induced by subharmonics in voices suggests that some essential factors in roughness perception were not considered in psychoacoustic models. A preliminary experiment was designed to demonstrate that no existing psychoacoustic theory could explain roughness induced by subharmonics.

9.3 Experimental study

9.3.1 Method

Two periodic complex tones S1 and S2 were synthesized by a PC computer with the same duration of 1 second and period $2/f_0 = 220$ Hz. Their spectra are given in Fig. 9.4. Both stimuli contain 15 harmonics nf_0 and 15 subharmonics $(2n\text{-}1)f_0/2$, for n ranging from 1 to 15. The harmonics envelope has a constant slope of 2 dB/rank. The subharmonics $(2n\text{-}1)f_0/2$ for $n > 2$ also have the same envelop slope but their amplitude is 10 dB smaller than their left adjacent harmonics.

Stimulus S1 differs from Stimulus S2 in having stronger components at $f_0/2$ and $3f_0/2$. The component at $f_0/2$ of Stimulus S1 is 12 dB stronger than the components at f_0, and its component at $3f_0/2$ follows the envelope slope of harmonics. In S2, the first subharmonic vanishes and the second subharmonic follows the envelop slope of subharmonics at higher frequencies.

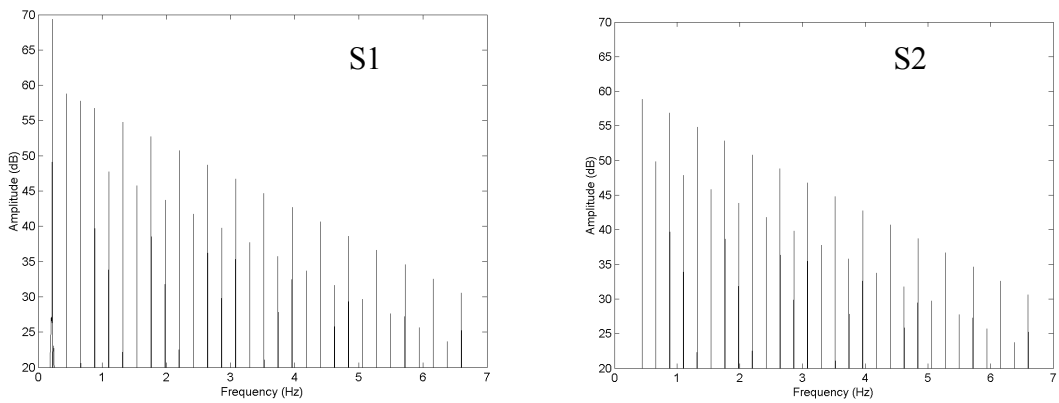


Figure 9.4: *Spectra of the two stimuli used in the perceptual experiment. Stimulus S1 differs from Stimulus S2 in having stronger components $f_0/2$ and $3f_0/2$.*

Subjects are 16 Taiwanese with ages ranging from 22 to 32 years. They are either students at the Conservatoire or amateur musicians. Stimuli S1 and S2 were presented with a 0.1 second pause. After listening to the two tones, subjects were asked to compare their impurity and pitch. The first question was two-alternative forced-choice: "which tone contains more impurities?" The second question for pitch comparison contained three options: (1) S1 and S2 are equal-pitch, (2) the pitch of S2 is an octave higher than S1, and (3) others.

9.3.2 Results and discussion

Experimental results showed that S2 has higher values of pitch and roughness than S1. For the task of roughness comparison, 13 of the 16 subjects reported that S2 contained more impurities than S1. For the task of pitch comparison, 14 of the 16 subjects reported that S2 was an octave higher than S1. The results showed a high inter-subject agreement.

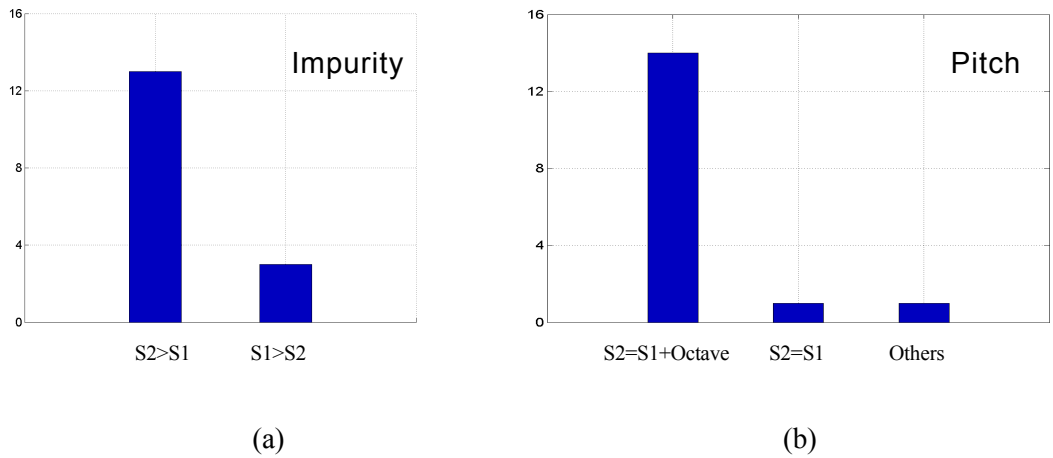


Figure 9.5: *Experimental results of roughness and pitch comparisons for S1 and S2.*

The fact that S2 sounds rougher than S1 contradicts past psychoacoustic models of roughness, because S2 was made from S1 by reducing the frequencies $f_0/2$ and $3f_0/2$. This experiment also indicates that roughness induced by subharmonics cannot be explained in terms of additional interference between components introduced by subharmonics. It has been suggested that when subharmonics exist in the source waveform, roughness is likely to be perceived because the fundamental and a subharmonic are in a critical band (Bergan and Titze 2001). This hypothesis ignores the fact that the lowest two subharmonics in speech and music are always weak and resolved by auditory channels; the increases in interference introduced by subharmonics are more important at higher frequencies than at lower frequencies, as high subharmonics and harmonics are unresolved by auditory channels. However, as S1 contains as many unresolved components as S2, according to existing models of roughness, they should have the same roughness. The present experimental results suggest that roughness increase is not due to the interference increases introduced by subharmonics. Note that the stimuli used here were strictly stationary. This is in sharp contrast to amplitude- or

frequency-modulated stimuli used in psychoacoustic and phoniatric research on roughness (Bergan and Titze 2001) but in line with a study of the "rustle" quality induced by stationary violin tones with subharmonics (Stepanek and Otcenasek 1999).

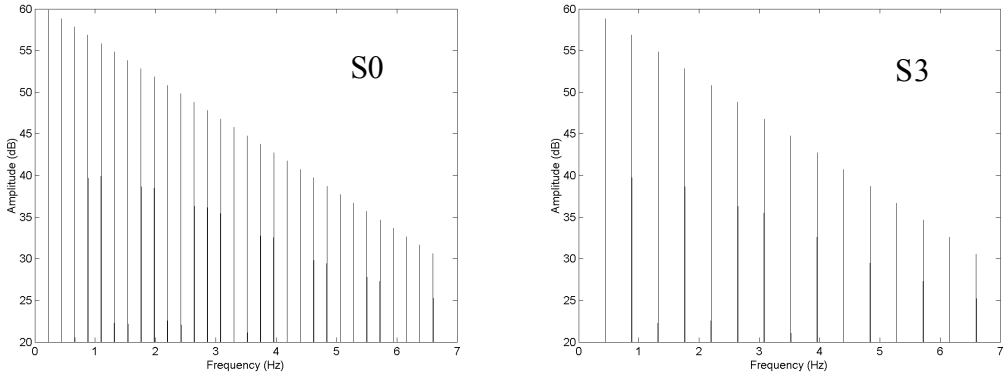
Roughness of the low-pitch tone S1 appears to be overestimated by psychoacoustic models of roughness. As mentioned in section 9.2.3, low-pitch natural tones do not sound rough, even when they contain rich unresolved components. The strong component $f_0/2$ in S1 may reduce interference among unresolved components at high frequencies through an integration mechanism across auditory channels. Note that S2 sounds an octave higher than S1; the strong component $f_0/2$ in S2 also contributes to the pitch sensation $f_0/2$. The correlation between low roughness and the strong pitch sensation of $f_0/2$ in S1 leads to a hypothesis that when the pitch strength of $f_0/2$ is high, interference induced by components $(2n\text{-}1)f_0/2$ is reduced. This assumption implies a mechanism integrating information across auditory channels. In the next section this mechanism will be identified as a grouping mechanism in auditory scene analysis.

The experimental result comparing pitches reveals the crucial role of the two lowest subharmonics $f_0/2$ and $3f_0/2$ in pitch extraction. For most periodic complex tones, the removal of the fundamental seldom affects their pitches. However, if the components $(2n\text{-}1)f_0/2$ are weaker than nf_0 – a spectral feature characteristic of natural tones containing subharmonics – the pitch is sensitive to the amplitude of the frequencies $f_0/2$ and $3f_0/2$.

When the subharmonic pitch $f_0/2$ competes with the pitch f_0, the task of rating their pitch strengths is not easily accomplished by means of the pitch models in frequency-domain, because there is no straightforward criterion for determining whether the harmonic templates of f_0 or $f_0/2$ fit the spectral lines better. The approach of autocorrelation analysis is more suitable for pitch strength estimations and comparisons.

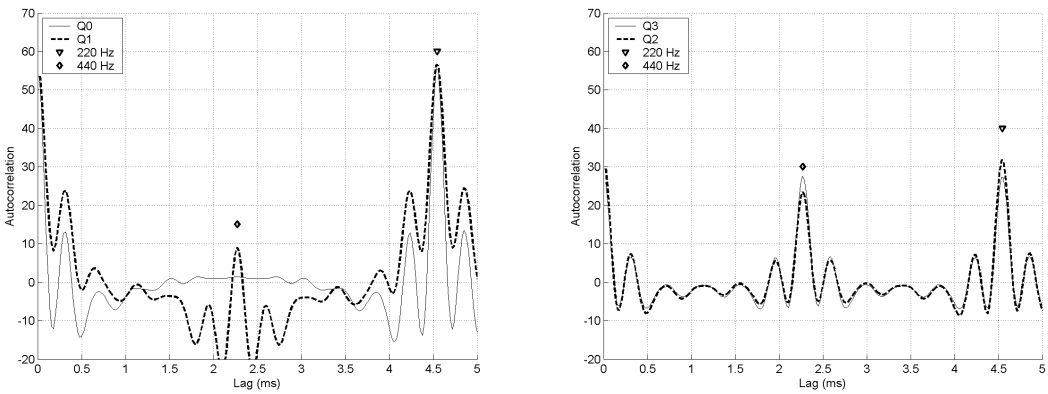
In order to demonstrate the performance of autocorrelation functions in pitch strength estimations, two subsidiary complex tones S0 and S3 were synthesized, whose spectra are given in Fig. 9.6. These two tones are complex tones with a constant spectral slope. Fig. 9.7 compares their autocorrelation functions with those of S1 and S2.[5] It can be observed that the autocorrelation functions of S0 and S1 are similar in having a peak at the pitch period $2/f_0$ [Fig. 9.7(a)]; hence, their pitches are both $f_0/2$. By contrast, the autocorrelation functions of S2 and S3 shown in Fig. 9.7(b) have peaks at the pitch periods $2/f_0$ (marked with '◊') and $1/f_0$ (marked with '▽'). Because the first peak corresponds to the pitch f_0, the pitch of S2 and S3 is f_0. This was borne out in the experimental result.

[5] Before autocorrelation analysis, these four sounds had been passed through a filter described in Appendix A.



(a) (b)

Figure 9.6: *Spectra of two complex tones with a constant spectral slope 2 dB/rank. (a) S0 with the fundamental frequency 220 Hz. (b) S3 with the fundamental frequency 440 Hz.*



(a) (b)

Figure 9.7: *Representation of pitch strengths of f_0 ('◊') and $f_0/2$ ('▽') in autocorrelation functions. (a) Comparison between the autocorrelation functions of the tones S0 and S1. The pitch $f_0/2$ = 220 Hz is more robust than f_0 = 440 Hz, as the first peak of their autocorrelation functions lies at $2/f_0$. (b) Comparison of the autocorrelation functions of the tones S3 and S2. The pitch f_0 = 440 Hz is more robust than $f_0/2$ = 220 Hz, as the first peak of their autocorrelation functions lies at $1/f_0$.*

9.4 Modeling roughness induced by subharmonics

9.4.1 Auditory scene analysis and grouping mechanisms

Roughness induced by subharmonics can be explained by taking into account grouping mechanisms in auditory scene analysis. The aspect of "auditory scene analysis" (Bregman 1990) deals with the organization of auditory scene which breaks a sound mixture into elements and 'unifies' proximate elements into discrete objects. As sounds rarely occur in isolation, auditory scene

analysis is important in the identification of sound objects. The analysis process consists of two conceptually distinct stages. In the first stage, sound is decomposed into a collection of sensory elements. Subsequently, elements that are likely to have arisen from the same acoustic source are grouped to form a perceptual whole, or 'stream'.

Group mechanisms in auditory scene analysis are considered to be governed by some 'grouping rules' such as common onset/offset, amplitude/frequency modulation, harmonicity and spatial location. Regarding the grouping rule of harmonicity, I assume a subsidiary rule for grouping: spectral components nf_0 are grouped only when the pitch sensation of f_0 is robust enough. In the other words, if the pitch strength of f_0 is low, components with frequencies nf_0 will not be grouped despite their harmonicity. This assumption highlights the ecological role of the pitch perception: when a lot of harmonic components reach the ears, a robust pitch sensation emerges in the auditory system *in order to* group them as a single entity.

Another assumption in my model is that interference between components lying in the same auditory channels will be largely reduced once these components are grouped. This assumption is suggested by the roughness of low-pitch male voices. As discussed in section 9.2.3, the values of auditory roughness in low-pitch male voices are generally overestimated by pyschoacoustic models of roughness. Interference between rich harmonics of low-pitch male voices may be reduced by the grouping mechanism according to pitch.

9.4.2 Model description

Based on these two assumptions, a qualitative model of roughness induced by subharmonics is proposed, which is schematically depicted in Fig. 9.8. This model consists of three steps: (1) pitch extraction, (2) sifting with a harmonic sieve, and (3) grouping and perception of harmonics/subharmonics.

9.4.2.1 Stage 1: Pitch extraction by autocorrelation analysis

The definition of subharmonics and harmonics demands a definition of pitch, which is not obvious when the subharmonic pitch $f_0/2$ competes with the pitch f_0. The subharmonic pitch is seldom perceived because of its low pitch strength. As subharmonics at low frequencies are much weaker than adjacent harmonics, they cannot evoke a robust subharmonic pitch. At high frequencies, subharmonics are comparable to adjacent harmonics in amplitude. However, their ranks are too high to contribute to the sensation of the subharmonics pitch. In Fig. 9.8(a), for instance, the subharmonics above $15f_0/2$ are noticeable. But their frequencies are much higher than the dominance region for pitch extraction, i.e. $3f_0/2$ to $5f_0/2$ (Ritsma 1967, Plomp 1967). Weakness of the subharmonic pitch can also be seen from its autocorrelation function [Fig. 9.8(b)], where the first peak lies at $1/f_0$ (marked with '▼'). Since the second peak at $2/f_0$ is only slightly higher than the first peak, the strength of the subharmonics pitch is fairly low.

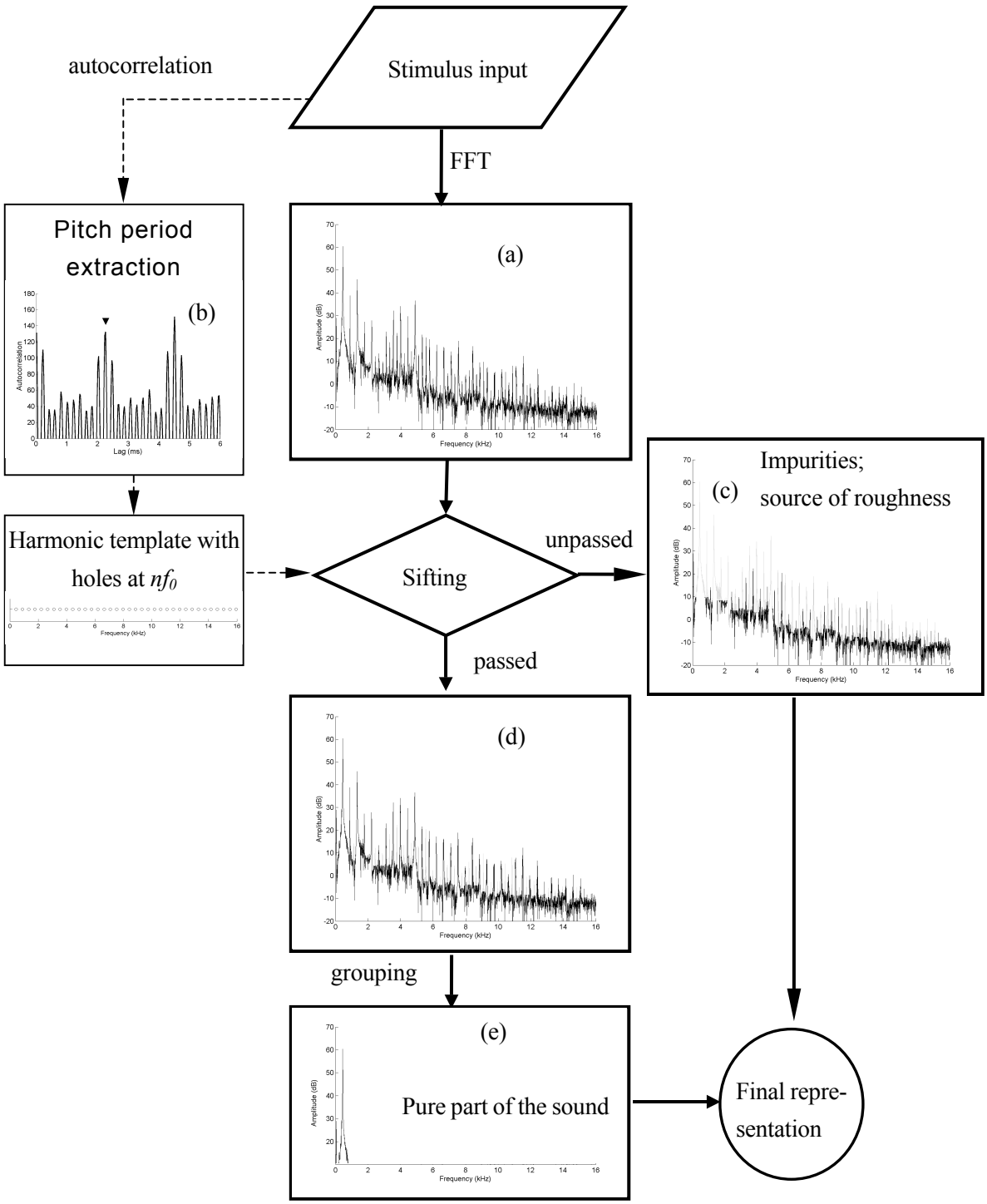


Figure 9.8: *Model of roughness induced by subharmonics. (a) Spectrum of a dizi tone. (b) Autocorrelation function. (c) Spectrum of subharmonics of the dizi tone. (d) Spectrum of harmonics of the dizi tone. (e) Harmonics grouped as a single auditory entity.*

9.4.2.2 Stage 2: Sifting with harmonic sieve

In this stage a harmonic sieve is constructed according to the pitch f_0. I follow recent models for *computational auditory scene analysis* to use autocorrelation analysis for f_0 extraction (Brown and Cooke 1994, Ellis 1996). This harmonic sieve consists of a series of harmonic "holes" at nf_0. Harmonics pass the sieve [Fig. 9.8(d)] while subharmonics are rejected by it [Fig. 9.8(c)].

9.4.2.3 Stage 3: Grouping and perception of harmonics/subharmonics

In the final step, the components passing the harmonic sieve are grouped as a single entity [Fig. 9.8(e)] and the unpleasant beatings between them are largely reduced. Therefore, the grouped harmonics of a rough *dizi* tone are not the main source of roughness. Rejected by the harmonic sieve, subharmonics remain ungrouped and evoke many entities in the auditory scene. They are perceived as the impurities of the sound.

9.4.3 Implications for future research

Although subharmonics are not grouped by a robust subharmonic pitch, they are not segregated from the sound object composed of grouped harmonics. Instead, subharmonics merely add impurities to this sound. This aspect appears to differ from the stream segregation in auditory scene analysis, where ungrouped components are separated from grouped components. In the perception of natural sounds with subharmonics, the auditory system still recognizes that both harmonics and subharmonics have arisen from the same source, possibly through a higher-order grouping based on learning. Since we often hear sounds with subharmonics emitted from one oscillator such as the glottis and musical instruments, the auditory system may have learned to bind subharmonics with harmonics.

9.5 Subharmonics in large *dizi* music

9.5.1 Subharmonic generation of the large *dizi*

The *dizi* tones in the third register are characterized by subharmonics. It is important to note that the generation of subharmonic depends on both the embouchure and the instrument. The amplitude of subharmonics in a *dizi* tone varies with the embouchure. A skillful musician can produce subharmonics-less *dizi* tones in the third register when rough tones are not desired. Nevertheless, there is a strong tendency for subharmonics to appear in the large *dizi*. It seems that musicians are not able to avoid producing subharmonics in soft tones in the third register when they play a large *dizi* with a slack membrane. Consequently, large *dizi* music is more characterized by a rough timbre than *qudi* or *bangdi* music.

A further reason for the importance of subharmonics in large *dizi* music is that the subharmonics in tones produced by smaller *dizi* result in a very annoying timbre because their spectral lines are

widened. It was reported in section 4.4.2 that impedance measurements on the *qudi* in C showed octave stretching for some fingerings. Because the first and second bore resonances for these fingerings do not stand in the ratio 1:2, when both modes are excited, the spectral lines of subharmonics are widened (see sections 3.4 and 6.4). The corresponding sound quality, distinguishable from roughness, is described by musicians as "sandy". This fairly unpleasant sound quality is generally avoided in *qudi* and *bangdi* music. In the large *dizi*, the spectral lines of subharmonics are not significantly widened, and the rough quality is appreciated in musical contexts when it is used appropriately.

9.5.2 Perception of subharmonics in large *dizi* music

Although rough large *dizi* tones with subharmonics always sound unpleasant because they contain impurities, they are optimal for imitating melancholic human voices. The appearance of subharmonics is a characteristic of pathological voices. But healthy humans with excessive emotional stress can also produce voices with subharmonics. For this reason rough singing has been used for emotional expressions in jazz music [see Fig. 9.9(a)] and Russian lament (Mazo *et al.* 1995).

The large *dizi* tones in the third register are optimal for imitating human voices with subharmonics, because their pitches are low enough to arouse an association with human singing. Examples in large *dizi* music imitating melancholic singing include "Weeping Flower" and "Double Rhyme Hatred". Subharmonics in *dizi* tones play a crucial role in producing the rough timbre of a weeping-like voice.

The large *dizi* has opened a new page in the history of *dizi* music, as it matches the traditional Chinese literati culture particularly well (Lin 1997). Some recent compositions for the large *dizi*, such as "A Wonderful Melody of Mountain *Langye*" and "Listening to Spring", show a philosophical impression similar to *Qin* music. In contrast to modern *qudi* and *bangdi* music, these three pieces are lightly accompanied and the tempo is slow. During a performance, the listener concentrates on nuances in timbre. Subharmonics as impurities can enrich the texture of large *dizi* tones.

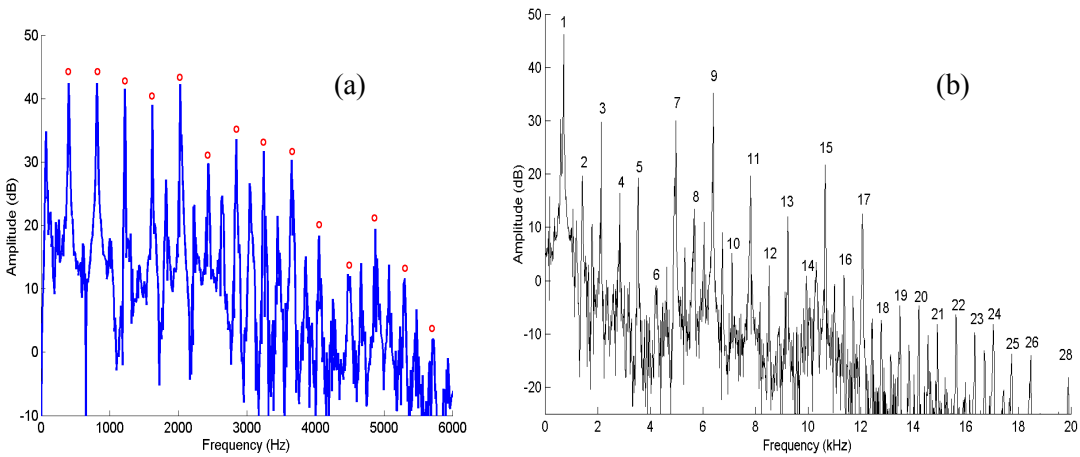


Figure 9.9: *Musical tones with subharmonics. (a) The saxophone. Spectral lines of harmonics are marked with* 'o'. *(b) The large dizi.*

For the *dizi* tones in the third register, roughness induced by subharmonics is sometimes combined with the multi-pitch effect due to the predominance of odd-numbered harmonics. Fig. 9.9 shows the spectrum of a large *dizi* tone, which is dominated by odd-numbered harmonics and contains subharmonics. As the 7th, 9th, and 11th, harmonics are much stronger than the adjacent even-numbered harmonics, the multi-pitches $2f_0'$ may be audible (see section 8.4). They combine with the impure subharmonics to produce multiple streams in the auditory scene.

9.6 Conclusions

Subharmonics with frequencies $(2n\text{-}1)f_0/2$ in the *dizi* tones in the third register induce a rough sound quality. Such rough sounds shed new light on auditory scene analysis by introducing the notion of *auditory impurity*; a sound with subharmonics is perceived as a "pure" part plus impurities. As impurities, subharmonics are segregated from the "pure" part of the sound composed of harmonics, but still bound to it through a higher-order grouping.

Subharmonics are more important in larger *dizi*. Because the tone range of a large *dizi* is comparable to human singing, it is especially optimal for imitating human voices. Subharmonics in large *dizi* tones often recall melancholic voices.

References

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit. Acustica 58, 268-281.

Arends, N., Povel, D.-J., van Os, E., and Speth, L. (1990). Predicting voice quality of deaf speakers on the basis of glottal characteristics. Journal of Speech and Hearing Research 33, 116-122.

Bergan, C.C., and Titze, I.R. (2001). Perception of pitch and roughness in vocal signals with subharmonics. Journal of Voice 15, 165-175.

Bregman, A.S. (1990). Auditory Scene Analysis. MIT Press.

Brown, G.J., and Cooke, M.P. (1994). Perceptual grouping of musical sounds: a computational model. Journal of New Music Research 23, 107–132.

Daniel, P., and Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. Acustica 83, 113-123.

Dejonckere, P.H., and Lebacq, J. (1983). An analysis of the diplophonia phenomenon. Speech Communication 2, 47-56.

Dejonckere, P.H. (1995). Principal components in voice pathology. Voice 4, 96–105.

Deutsch, D. (1999). Grouping mechanism in music. In: The Psychology of Music (edited by D. Deutsch), Academic Press.

Feigenbaum, M.J. (1978). Quantitative universality for a class of nonlinear transformation. Journal of Statistic Physics 19, 25-35.

Eggermont, J.J. (1998). Representation of spectral and temporal sound features in three cortical fields of the cat: Similarities outweigh differences. Journal of Neurophysiology 80(5), 2743-2764.

Ellis, D.P.W. (1996). Prediction-driven computational auditory scene analysis. PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Fishman, Y., Reser, D.H., Arezzo, J.C. and Steinschneider, M. (2000). Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. Journal of the Acoustical Society of America 108(1), 235-246.

Gaese, B.H., and Ostwald, J. (1995). Temporal coding of amplitude and frequency modulation in the rat auditory cortex. European Journal of Neurosciences 7, 438-450.

Gibiat, V., and Castellengo, M. (2000). Period doubling occurrences in wind instruments musical performance. Acustica 86, 746-754.

Gomperts, A. (1995). Tunings, Tone Systems and Psychoacoustics of Sundanese, Javanese and Balinese Music. In: Oideion: The Performing Arts World-Wide (edited by W. van Zanten and M. van Roon), vol. 2: Ethnomusicology in the Netherlands; Present Situation and Traces of the Past (= CNWS Publications vol. 35), Leiden: Research School CNWS, S. 173-207.

Helmholtz, H.L.F. (1877). On the Sensations of Tone, 4th ed., translated by A.J. Ellis, Dover, New York, 1954.

Herzel, H., Berry, D.A., Titze, I.R., and Saleh, M. (1994), Analysis of vocal disorders with methods from nonlinear dynamics, Journal of Speech and Hearing Research 37, 1008-1019.

Herzel, H., Berry, D., Titze, I.R., and Steinecke, L. (1995). Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. Chaos 5, 30-34.

Kimura, M. (1999). How to produce subharmonics on the violin? Journal of New Music Research 28(2), 178-184.

Lauterborn, G.M. (1996). The bowed string as a nonlinear dynamical system. Acustica 82, 657-664.

Langer, G. (1992). Periodicity coding in the auditory system. Hearing Research 60, 115-142.

Leman, M. (2000). Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM). In: Proceeding of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona.

Mazo, M. Ericson, D., and Harvery, T. (1995). Emotion and expression: Temporal data on voice quality in Russian lament. In: Vocal Fold Physiology: Voice Quality Control (editede by O Fujimura and M. Hirano), Singular, San Diego, pp. 173-178.

McKinney, M.F. (2001). Neural correlates of pitch and roughness: Toward the neural code for melody and harmony perception. PhD Diss., Massachusetts Institute of Technology, Cambridge, MA.

McKinney, M.F., Tramo, M.J., and Delgutte, B. (2001). Neural correlates of musical dissonance in the inferior colliculus. In: Physiological and Psychophysical Bases of Auditory Function (edited by D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven), Shaker Publishing BV, 83-89.

Mergell, P., Herzel, H., and Titze, I. (2000). Irregular vocal fold vibration – High-speed observation and modeling. Journal of the Acoustical Society of America 108, 2996-3002.

Møller, A.R. (2000). Hearing: its Physiology and Pathophysiology. Academic Press.

Omori, K., Kojima, H., Kakani, R., Slavit, D.H., and Blaugrund, S.M. (1997). Acoustic characteristics of rough voice: Subharmonics. Journal of Voice 11(1), 40-47.

Plomp, R., and Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. Journal of the Acoustical Society of America 38, 548-560.

Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (2000). Perception of musical tension for non-tonal orchestral timbres and its relation to psychoacoustic roughness. Perception and Psychophysics, 62(1), 66-80.

Reuter, R. (2000). Untersuchung der Rauhigkeit menschlicher Stimmen auf der Grundlage der nichtlinearen Dynamik und der Psychoakusik. PhD thesis, TU Berlin.

Riede, T., Wilden, I., and Tembrock, G. (1997). Subharmonics, biphonations, and frequency jumps – common components of mammalian vocalization or indicators for disorders. Z. Säugetierkunde 62, Suppl. 2, 198-203.

Steinecke, I., and Herzel, H. (1995). Bifurcations in an asymmetric vocal fold model. Journal of the Acoustical Society of America 97, 1571-1578.

Stepanek, J., and Otcenasek, Z. (1999). Rustle as an Attribute of Timbre of Stationary Violin Tones, Journal of the Catgut Acoustical Society 3(8), 32-38.

Sun, X., and Xu, Y. (2002). Perceived pitch of synthesized voice with alternate cycles. Journal of Voice, 16(4), 443-459.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). Acustica 30, 201–212.

Tramo, M.J., Cariani, P.A. and Delgutte, B. (1992). Representation of tonal consonance in the temporal firing patterns of auditory-nerve fibers: Responses to musical intervals composed of pure tones vs. harmonic complex tones. Society of Neurosciences Abstracts.

Tramo, M.J., Cariani, P.A., McKinney, M.F. and Delgutte, B. (2000). Neural coding of tonal consonance and dissonance. Abstracts of the 2000 Midwinter Meeting of the Association for Research in Otolaryngology.

Tramo, M.J., Cariani, P.A., Delgutte, B., and Braida, L.D. (2001). Neurobiological foundations for the theory of harmony in western tonal music. In: The Biological Foundations of Music (edited by R.J. Zatorre and I. Peretz), The New York Academy of Sciences, 92-116.

Zwicker, E., and Fastl, H. (1990). Psychoacoustics: Facts and Models. Springer-Verlag.

Chapter 10.
Conclusions

10.1 Physics of the *dizi*

The object of this thesis was the Chinese membrane flute, the *dizi*, which has a membrane covering a hole in the wall of its resonator. Studies in the mechanics of the *dizi* have explained the importance of the tension and wrinkles in the membrane in producing a beautiful, bright *dizi* timbre, which enables the *dizi* to 'cut through' the sound of even large ensembles. Investigations on the linear and non-linear properties of the *dizi* have provided insights into the physics of this membrane flute: the membrane gives a rich spectral content to the flute at the expense of pitch range.

10.1.1 Impedance correlates of *dizi* tone range

The linear behavior of the *dizi* was studied by modeling its membrane as a piston. Impedance measurements gave typical values of parameters in the piston model: the resonant frequency 2 kHz < f_m < 3 kHz, the effective acoustic mass m of the order of 10^{-6} kg, and the damping coefficient R of the order of 10^{-3} kg/sec.

Input impedance calculations of the *dizi* resonator, a pipe-membrane system, showed some phenomena common in soundboard-string systems. In the case of the membrane located at the middle of the pipe, it does not affect the even-numbered bore resonances, as it is located at pressure nodes of the corresponding standing waves. On the other hand, the odd-numbered bore resonances are shifted away from f_m and their admittances are reduced by the membrane, as it is located at a pressure antinode of the corresponding standing waves (see Fig. 4.5). Admittance reductions were found to be significant for the bore resonances close to the membrane's resonant frequency f_m.

Resonance shifts and admittance reductions predicted by the linear piston model quantitatively agreed with the results of experiments on a simple pipe-membrane system at low frequencies (< f_m). Impedance measurements of the *qudi* in C showed resonance shifts caused by the membrane of up to 60 cents. It was also found that admittances of its bore resonances of the *qudi* in C supporting the notes B6 and C7 were largely reduced by the membrane. Since the vibration modes of the pipe-membrane resonator are damped by the membrane for resonances close to f_m, the typical *dizi*

tone range is restricted to two octaves plus two notes.

10.1.2 Membrane-wrinkling correlates of *dizi* timbre and playability

The nonlinear behavior and wrinkling of a *dizi* membrane, essential to production of the characteristic *dizi* timbre (Thrasher 2001), were studied theoretically and experimentally. The motion of the wrinkled membrane can be described as a series of strings, whose behavior is fairly linear until the tension increases for large deformations. Theoretical calculation suggested that a wrinkled membrane has a higher initial tension ($4N_0$) than a membrane with an isotropic initial tension (N_0) for the same resonant frequency. Wrinkling reduces the cubic non-linearity of the membrane by decoupling the strings and allowing them to have a high initial tension.

The existence of the membrane's non-linearity was demonstrated by driving it with sinusoidal acoustic waves strong enough to induce its harmonic generation. Relations between the magnitudes of the lowest three harmonics revealed a predominance of cubic non-linearity for an unwrinkled membrane, and a predominance of quadric non-linearity for a wrinkled membrane in an intensity range of the driving signals. The membrane's response curves showed the frequency-pulling effect characteristic of a hardening spring. A jump phenomenon was observed in the multi-valued response curve of a slack, unwrinkled membrane. In an impedance calculation, the membrane with a large/small jump was approximated as a linear oscillator with a large/small damping coefficient. The large damping could lead to admittance reductions of the bore resonances supporting notes for which the membrane is located near a pressure antinode of the fundamental.

The main effect of wrinkling is to make the membrane behave more linearly by reducing the cubic non-linearity. If the membrane is unwrinkled due to an isotropic initial tension, the spring constant will be too high for a high initial strain so that the membrane hardly vibrates, or, the cubic non-linearity will be too high for a low initial strain. In the latter case, a jump phenomenon in the response curve of the membrane can lead to hysteresis and discontinuous behavior of the *dizi*: for the same fingering it either does not sound, or sounds raucously. Wrinkling largely stabilizes its sound production.

10.1.3 Non-linear mechanism correlates of spectral features of *dizi* tones

The *dizi* membrane was modeled as a Duffing oscillator driven by acoustic waves in the tube. This model successfully explained three major spectral features of *dizi* tones: (1) first formant, (2) predominance of odd-numbered harmonics, and (3) subharmonics at high frequencies. It was tested by two kinds of quasi-sinusoidal tones driving the membrane: tones generated by external excitation or by blowing the instrument. The *dizi* tones for which the membrane is located near a pressure node of the second harmonic wave were simulated with the motion equation of the membrane:

$$\ddot{x} + r\dot{x} + w_m^{\,2}x + bx^3 = \frac{G_0}{m}\cos(wt + j_0) + \frac{G_1}{m}\cos(2wt + j_1) + \frac{G_2}{m}\cos(wt/2 + j_2) \qquad (10.1)$$

where

$$b > 0, \quad G_0 >> G_1 \text{ and } G_0 >> G_2 .$$

For external excitation, the phase plots of the membrane showed interlocking spirals, which were predicted by this quasi-sinusoidally driven Duffing oscillator model. For normal *dizi* tones, the phase plots of the membrane also showed similar interlocking spirals, but rich harmonics above 10 kHz in *dizi* tones did not appear in simulated spectra.

The Duffing oscillator model of the *dizi* membrane can explain why the spectral features of *dizi* tones vary with pitch. According to spectral characteristics, the *dizi* tone range is divided into five registers. The upper notes of the first octave are characterized by a predominance of odd-numbered harmonics, and therefore the first octave is divided into two registers. In the first register, the membrane is located near a pressure antinode of the acoustic wave of the second harmonic, thus driven by both the fundamental and the second harmonic. That is why the radiating sounds contain both odd-numbered and even-numbered harmonics. On the other hand, for the *dizi* tones in the second register, the membrane is located at a pressure node of the second harmonic wave. Because the force driving the membrane is quasi-sinusoidal, this Duffing oscillator produces tones with dominant odd-numbered harmonics.

In the third register (lower notes of the second octave), in addition to a predominance of odd-numbered harmonics, subharmonics $(2n\text{-}1)f_0/2$ may appear. Subharmonics at high frequencies stem from the subharmonic force $G_2\cos(\omega t/2)$ exerted on the membrane, which is due to the excitement of the first mode of the air column. For a membrane-less flute, the subharmonics $(2n\text{-}1)f_0/2$ are restricted to low frequencies. For a membrane flute, rich high-frequency subharmonics are generated by the non-linear membrane under the weak subharmonic force $G_2\cos(\omega t/2)$ combined with the strong fundamental force $G_0\cos(\omega t)$.

The fourth register (upper notes of the second octave) is characterized by the absence of high harmonics because the membrane is located near a pressure node of the fundamental, whereas the fifth register (lower notes of the third octave) is characterized by a dramatic recovery of high harmonics because the membrane is located near a pressure antinode of the fundamental.

Dizi tones in the lowest three registers always have two formants, typically centered at 4–6 kHz and 10–14 kHz. Numerical simulations successfully predicted that the first formant frequency increases with the tone intensity, but failed to explain the harmonics above 10 kHz. The second formant reveals the filtering effect of the lowest transverse pipe mode with the resonance estimated as $c/2D$ (D is the pipe's diameter).

10.2 Perception of *dizi* tones

The timbre of *dizi* has been studied in terms of its static spectral features. The global pattern of spectral envelopes of *dizi* tones, represented as the auditory attribute "brightness", has been

distinguished from envelope jaggedness due to a predominance of odd-numbered harmonics, which has been related to the hollow and nasal qualities described by Helmholtz. In addition to harmonics, subharmonics have been found to play a role in the perception of *dizi* tones.

10.2.1 Brightness in *dizi* music

The spectral distribution of harmonics of *dizi* tones can extend to 22 kHz, and their auditory attributes vary with the frequency region. The harmonics around the first formant, responsible for brightness, can articulate stream segregation of the *dizi*. The membrane raises loudness of flute tones by transferring acoustic energy from lower frequencies to higher frequencies, to which our auditory system is more sensitive. On the other hand, the harmonics around and above the second formant can reduce breathiness by masking the turbulence noises.

The bright quality of *dizi* tones introduced by the membrane is not uniform in the entire tone range. *Dizi* tones in the fourth register lose the rich spectral content characteristic of a membrane flute. The inhomogeneity of brightness induces some interesting auditory associations or illusions. Lin (1997) mentioned the spatial effects of abrupt brightness changes in *dizi* music, i.e. when a melody ascended from bright notes to dull notes, the sound source seemed to move to a remote place (distancing effect) and become more difficult to localize (localization-confusion effect).

The distancing effect can be explained in terms of spectral cues for auditory distance perception. Since a sound loses more high frequencies while traveling through air, unexpected absence of high-frequency harmonics of *dizi* tones might be interpreted by the auditory system as an increasing distance from the *dizi*.

The localization-confusion effect can be explained in terms of monaural spectral cues for sound source localization. A lack of spectral components above 4 kHz is likely to cause a confusion of sound elevation that is estimated according to high-frequency monaural cues.

Because of the involvement of other cues for auditory distance perception such as loudness, these two spatial effects could not actually change the *psychophysical* distance of the sound source. Yet, they may affect the *aesthetic* image of the *dizi* in music. The *dizi* with an inhomogeneous timbre is thus optimal for creating a multi-level auditory impression without changing the physical position of the instrument.

10.2.2 Auditory correlates of the predominance of odd-numbered harmonics

Dizi tones in the second and third registers are always characterized by the predominance of odd-numbered harmonics, which is closely related to the hollow and nasal qualities described by Helmholtz. I found that nasal voices produced with a water membrane in the nasal cavity have the spectral feature of a predominance of odd-numbered harmonics around 5 kHz and a sounding mechanism similar to those of the *dizi* tones.

An analogy was proposed, comparing the spectral feature of large amplitude differences between odd- and even-numbered harmonics and the visual pattern of luminance-contrasting gratings.

As these gratings make a strong visual impression, so too does a predominance of odd-numbered harmonics evoke a distinct timbre. When this predominance occurs at high frequencies, interference between adjacent harmonics are largely reduced. This may explain why nasal *dizi* tones are described as 'sweet' by musicians.

The predominance of upper odd-numbered harmonics can induce additional pitches, which enrich the texture in *dizi* solo music. An autocorrelation model of pitch suggested two necessary conditions for the existence of the additional pitches: (1) odd-numbered harmonics must predominate in the frequency range above the 5th harmonics, and (2) the average amplitude difference between odd- and even-numbered harmonics in this range must be larger than 8 dB. Since these two conditions are always satisfied in the *dizi* tones in the second and third registers but not in most tones produced by other musical instruments, this multi-pitch phenomenon is fairly unique to the *dizi*.

10.2.3 Roughness induced by subharmonics

Subharmonics at $(2n\text{-}1)f_0/2$ sometimes appear in the *dizi* tones in the second octave. These subharmonics do not produce another pitch one octave below, but merely induce a rough sound quality. Roughness induced by subharmonics cannot be explained by past psychoacoustic models and appears against the theory of consonance/dissonance.

Roughness induced by subharmonics can be explained by taking into account a grouping mechanism that integrates information across auditory channels. A qualitative model was proposed with two assumptions: (1) grouping components at nf_0 demands a robust pitch sensation of f_0; (2) unpleasant beatings caused by components in the same critical bands are largely reduced by this grouping. Since subharmonics in *dizi* tones at low frequencies are always much weaker than adjacent harmonics, the subharmonic pitch of $f_0/2$ cannot be evoked. Therefore, the ungrouped subharmonics are perceived as impurities of *dizi* tones.

Subharmonics are more important in larger *dizi*. A large *dizi* with a thin membrane has a strong tendency to produce upper subharmonics. Because the tone range of a large *dizi* is comparable to human singing, it is especially optimal for imitating human voices. Large *dizi* tones with subharmonics often recall melancholic voices.

10.3 Perspectives

10.3.1 Aerodynamics

As the non-linearity of the membrane cannot explain the spectral content of *dizi* tones above 10 kHz, another non-linear mechanism in this instrument, i.e. the aerodynamics at the mouth hole, should be taken into consideration. If the jet amplifies the harmonics generated by the membrane in the range 3–5 kHz, they may in turn drive the membrane to radiate harmonics above 10 kHz. Two mechanisms relevant to the jet's sensitivity to high-frequency harmonics were proposed: the moving

separating points of the jet, and its varicose oscillations due to the pressure fluctuations at the flue exit. For a membrane-less flute, high-frequency harmonics are so weak that these effects could be neglected. For the *dizi*, the non-linear membrane is able to produce high harmonics, bringing the question of the high-frequency response of the jet into prominence. This invites further physical studies on membrane flutes.

10.3.2 Biomusicology

Dizi tones have several interesting spectral features, such as the brightness inhomogeneity, the predominance of odd-numbered around 4–6 kHz, and the subharmonics at high frequencies. Auditory effects associated with these features have shed new light on the relationship between the ecological role of auditory processing and music perception.

While the new field "biomusicology" places the analysis of music origin and its application to the study of human origins at its very foundation (Brown *et al.* 2001), *dizi* music demonstrates *limitations* of the auditory system. For instance, abrupt brightness changes in *dizi* music induce spatial effects by exploiting the limited elasticity of the auditory system and the assumption of brightness constancy. *Dizi* music also demonstrates performance limits of the pitch-based grouping mechanism in auditory scene analysis. A *dizi* tone with subharmonics elicits the pitch which is *too high* to group the subharmonics. On the other hand, a *dizi* tone with the predominance of odd-numbered harmonics elicits *too many* pitches for grouping. Future research might be dedicated to discovering the neural representations of the spatial effects and stream segregation/grouping in auditory scene analysis.

References

Brown, S., Merker, B., and Wallin, N.L. (2001). An introduction to Evolutionary Musicology. In: The Origins of Music (edited by N.L. Wallin, B. Merker and S. Brown), MIT Press, 3-24.

Cerda, E., and Mahadevan, L. (2003). Geometry and physics of wrinkling. Physical Review Letters 90, 074302.

Hagedorn, P. (1995). Mechanical oscillations. In: Mechanics of Musical Instruments (edited by A. Hirschberg, J. Kergomard and G. Weinreich), Springer-Verlag, 7-78.

Heffner, R.S., and Heffner, H.E. (1992). Evolution of sound localization in mammals. In: The Evolutionary Biology of Hearing (edited by D.B. Webster, R.R. Fay and A.N. Popper), Springer-Verlag, 691-716.

Lin, G.F. (1997). Dikuan Yioqin. （林谷芳《諦觀有情：中國音樂裡的人文世界》，台北：望月文化）

Thrasher, A. (2001). "*Di*" in The New Grove Dictionary (edited by S. Sadie), Macmillan, Vol. 7, 277-279.

.

Appendix A
Autocorrelogram for multi-pitch transcription

The method of representing multi-pitches of quasi-periodic tones in autocorrelogram was developed for *dizi* solo music transcription. The program used is based on several autocorrelation models of pitch extraction (Meddis and Hewitt 1991, Cariani and Delgutte 1996, Meddis and O'Mard 1997). This appendix describes the principles and applications of this Matlab-implemented program.

A.1 Method description

The autocorrelogram computation for an input signal contains three stages: (1) band-pass filtering, (2) calculating the autocorrelation function in a moving time-window, and (3) plotting the value of the running autocorrelation function.

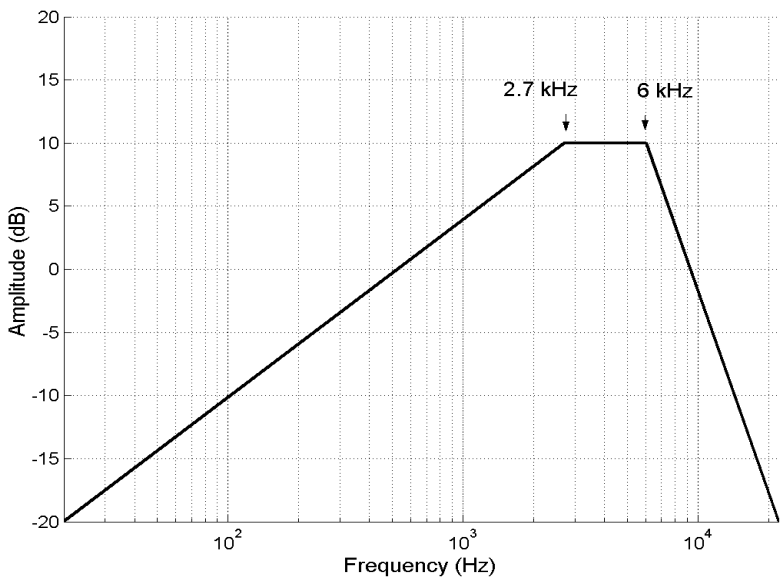


Figure A.1: *Band-pass filter used before autocorrelation analysis.*

The band-pass filtering is described in Fig. A.1. Physiologically, it corresponds to the outer/middle ear filtering considered in Meddis and Hewitt's (1991) model and the octave band weights according to the proportion of fibers in a human characteristic frequency distribution considered in Cariani and Delgutte's (1996) model. Psychoacoustically, this filtering is consistent with the equal-loudness contours (Glasberg and Moore 1990). However, this filter used here is highly simplified and has a steeper attenuation at low frequency than those in previous models. This filter, aimed to amplify the harmonics around the first formant, demands further research for a justification.

After signal filtering, autocorrelation is calculated in a finite-duration time-window, which shifts along the time axis. The running autocorrelation of a digital signal *wav*(*t*) at time *t* and lag τ is given by

$$autocorrelation(t, \tau) = \sum_i wav(t + i\Delta t) wav(t + i\Delta t + \tau) \qquad \text{(A.1)}$$

where Δt is the sample period and the summation is taken for *i* ranging from 1 to $W/\Delta t$ with *W* the width of the time-window. Their values in the present program are Δt = 1s/44100 = 0.023 ms and $W = 256\Delta t$ = 5.8 ms. It is important to note that in autocorrelation analysis I follow Karjalainen and Tolonen (1999) to skip the cochlear filtering for reducing the heavy computation of filterbank in Meddis and Hewitt's (1991) model. This simplification also requires further studies for a justification.

In plotting the values of autocorrelation functions, plot grayness is determined by

$$\frac{autocorrelation(t, \tau)}{MAX[autocorrelation(t, *)]} - 0.6 \qquad \text{(A.2)}$$

The maximum of the autocorrelation function at time *t*, represented at the prominent pitch period in the autocorrelation function, is used for normalizing the autocorrelation. It is assumed here that peaks in the autocorrelation function are of perceptual importance only when their heights are larger than 0.6 times the major peak height. The *y*-axis of an autocorrelogram is in the cent scale for fitting the lines of the soprano clef (see Figs. 8.10 and 8.11). The musical pitches are extracted by observing the peaks of the running autocorrelation, as the following examples will demonstrate.

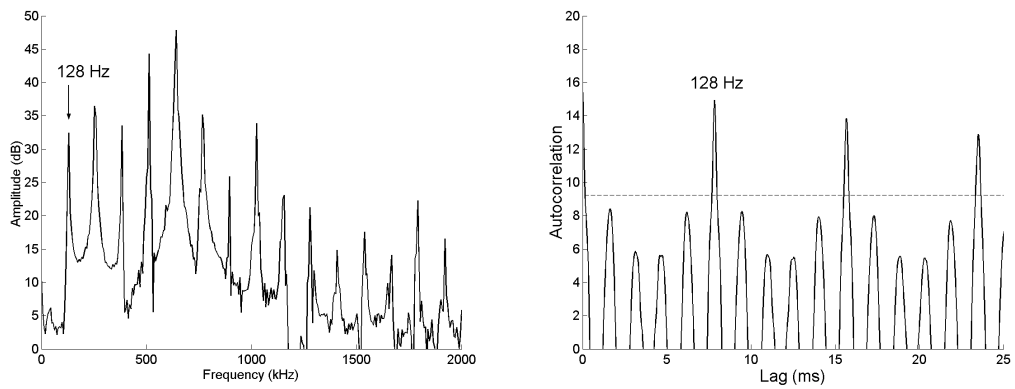
A.2 Examples of autocorrelation functions

A.2.1 Single-formant vowel

Fig. A.2 gives the spectrum and the autocorrelation function of a single-formant vowel after

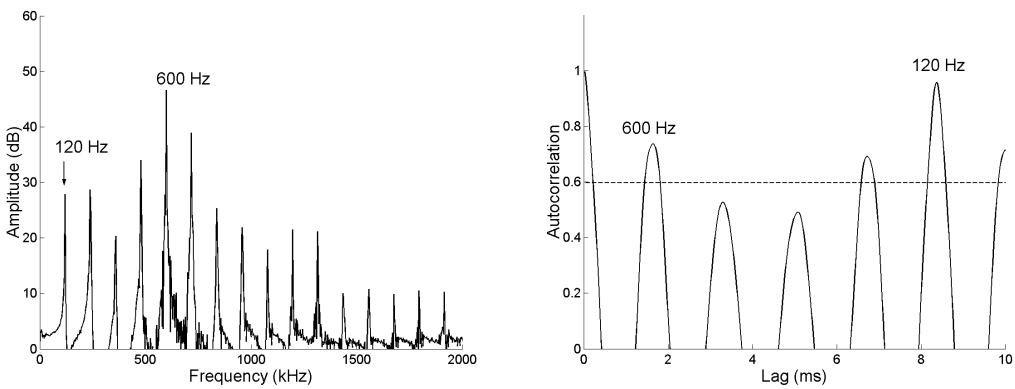
being passed through the filter described in Fig. A.1. The fundamental frequency 128 Hz is represented in the autocorrelation function as the first major peak at its period. Note that the autocorrelation function shows repetitive patterns beyond this period. That is why only the first major peak can represent the pitch.

The formant of this vowel is represented as minor peaks dividing the major peaks. Note that their heights are below the threshold for pitch sensation, which is assumed to be 0.6 times the major peak height (dash line). So this sound has only one pitch.



(a) (b)

Figure A..2: *Representations of pitch of a single-formant vowel. (a) Spectrum of the voice after being passed through the filter described in Fig. A.1. (b) Autocorrelation function. The fundamental frequency 128 Hz is represented as the first major peak at its period.*



(a) (b)

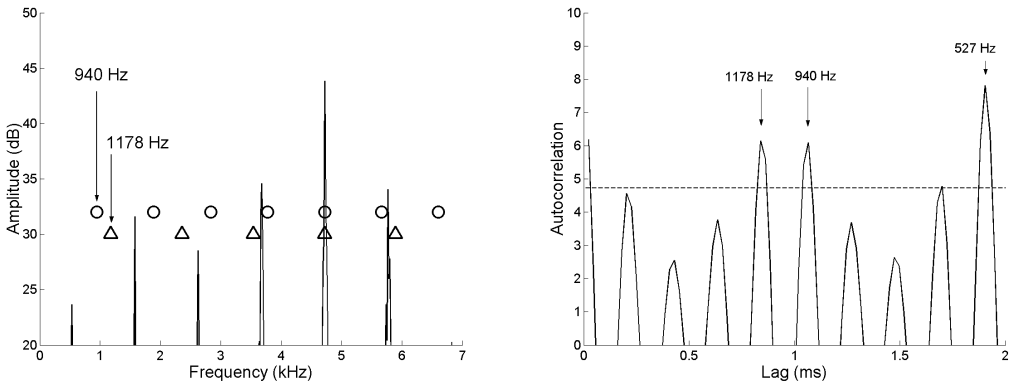
Figure A.3: *Representations of pitch of a vowel with a strong formant centered at 600 Hz. (a) Spectrum of the voice after being passed through the filter described in Fig. A.1. (b) Autocorrelation function. The minor peaks below the major peak represent the formant. As the first one of them is beyond the threshold for pitch sensation (dash line), it evokes a second pitch of 600 Hz beside the fundamental frequency 120 Hz.*

A.2.2 Stream segregation of a strong harmonic in a vowel

When the formant of a vowel is strong enough, it will segregate from the prominent pitch of the fundamental frequency and evoke a second pitch. Fig. A.3 gives the spectrum and the autocorrelation function of a vowel with a strong formant. Note that the first of the minor peaks representing the formant is beyond the threshold for pitch sensation (dash line). So this sound has two pitches: the fundamental frequency 120 Hz and the frequency of the fifth harmonic 600 Hz. The stream segregation of a strong harmonic in a vowel finds its practical importance in *overtone singing* (Adachi and Yamada 1999).

A.2.3 Multi-pitches due to the odd-numbered harmonics predominance

Some *dizi* tones can evoke multi-pitches that are perceived as segregated streams in auditory scene: listeners hear "something" above the prominent pitch. Multi-pitches in *dizi* music can be explained by temporal and place models of pitch extraction.



(a) (b)

Figure A.4: *Representations of multi-pitches of a dizi tone with dominant odd-numbered harmonics. (a) Spectrum of the dizi tone and the harmonic templates of 940 Hz (marked with '○') and 1178 Hz (marked with '△'). (b) Autocorrelation function of the dizi tone after being passed through the filter described in Fig. A.1. The periods of pitches are represented as peaks in the autocorrelation function.*

Consider a *dizi* tone with dominant odd-numbered harmonics [Fig. A.4(a)]. Since the spectral lines around the first formant are 'double-spaced', weak pitches approximately an octave above the fundamental frequency 527 Hz could be evoked. According to a pitch model in the frequency domain (Goldstein 1973, Terhardt 1974, Terhardt 1982), the strong 7th, 9th and 11th odd-numbered harmonics could induce two pitches of 940 Hz and 1178 Hz, because their harmonic templates fit these three odd-numbered harmonics. In the time domain, these two pitches are also represented in the autocorrelation function, which shows two peaks at their periods [Fig. A.4(b)].

The pitch perception correlated to the odd-numbered harmonics predominance is supported by psychoacoustic studies using alternative-polarity clicks (Flanagan and Gutman 1964) and stimuli composed of equally spaced inharmonic components (Schouten *et al.* 1962, Patterson 1973, Patterson and Wightman 1976).

A.3 Matlab implementation

```
% autocorrelogram for multi-pitch transcription of dizi solo music
%
%   paramaters
%   wav           input digital signal
%   fs            sample rate of input digital signal
%   ln            length of input digital signal
%   duration      duration of input digital signal
%   window        width of moving time-window in samples
%   t_resolution  step of moving time-window in samples
%   lag_m         time-lag for autocorrelation
%   sm2           value of autocorrelation function
%   y0            maximum of autocorrelation function
%   p_st          normalized value of autocorrelation function
%   gg            plot grayness


clear;
hold on;
[wav,fs]=wavread('mp0_10.wav');
ln=length(wav);
duration=ln/fs;

t_resolution=512;
window=256;
lag_m=150;

for tt=1:round(ln/t_resolution),
    t=(tt-1)*t_resolution;
    for lag=1:lag_m,
        for k=t+1:t+window,
            sm1(k)=wav(k)*wav(lag+k);
        end
        sm2(lag)=sum(sm1(t+1:t+window));
    end

    [y0,i0]=max(sm2(20:150));

    if y0>0.001
        for pp=20:150,
            p_st=sm2(pp)/y0-.6;
            if p_st<0
                gg=1;
            else
                gg=1-p_st*2;
            end
            x1=t/fs;
```

```
            x2=(t+t_resolution)/fs;
            y1=log10(fs/pp)/log10(2)*1200-9930;
            y2=log10(fs/(pp+1))/log10(2)*1200-9930;
            patch ([x1 x2 x2 x1],[y1 y1 y2 y2],[gg gg 1]);
        end
    else
    end
end

lines=[0 4 7 11 14 17 21 24 28 31 35]*100;
for nl=1:11,
    plot ([0 duration],[1 1]*lines(nl),'k');
end

 axis([0 duration 0 3500]);
 xlabel('Time (sec)');
 ylabel('Cent');
```

References

Adachi, S., and Yamada, M. (1999). An acoustical study of sound production in biphonic singing, Xöömij. Journal of the Acoustical Society of America 105(5), 2920-2932.

Cariani, P.A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. Journal of Neuro-physiology 76, 1698–1716.

de Boer, E. (1956). On the “Residue” in Hearing. Unpublished PhD thesis, University of Amsterdam.

Flanagan, J.L., and Gutman, N. (1964). Pitch of high-pass-filtered pulse trains. Journal of the Acoustical Society of America 36, 757-765.

Glasberg, B.R., and Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. Hearing Research 47, 103–138.

Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. Journal of the Acoustical Society of America 54, 1496-1516.

Karjalainen, M., and Tolonen, T. (1999). Multi-pitch and periodicity analysis model for sound separation and auditory scene analysis. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP’99), vol. 2, 929-932.

Licklider, J.C.R. (1951). A duplex theory of pitch perception. Experientia 7(4), 128-134.

Meddis, R., and Hewitt, M.J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. Journal of the Acoustical Society of America 89, 2866–2882.

Meddis, R., and O’Mard L. (1997). A unitary model of pitch perception. Journal of the Acoustical Society of America 102, 1811–1820.

Patterson, R.D. (1976). The effects of relative phase and the number of components on residue pitch. Journal of the Acoustical Society of America 53, 1565-1572.

Patterson, R.D., and Wightman, F.L. (1976). Residue pitch as a function of component spacing. Journal of the Acoustical Society of America 59, 1450-1459.

Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962). Pitch of the residue. Journal of the Acoustical Society of America 34, 1418-1424.

Terhardt, E. (1974). Pitch, consonance, and harmony. Journal of the Acoustical Society of America 55, 1061-1069.

Terhardt, E., Stoll, G., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. Journal of the Acoustical Society of America 71(3), 679-688.

Curriculum Vitae

Chen-Gia Tsai（蔡振家）

6 April 1971	born in Yünlin (Taiwan)
1990–1994	Bechlor of Science, Department of Physics, National Taiwan University, Taipei, Taiwan
1994–1997	Master of Arts, Graduate School of Folk Culture and Arts, National Institue of the Arts, Taipei, Taiwan
January–February 2002	Visiting student, School of Physics, University of New South Wales, Sydney, Australia
2000–2003	Ph.D. *am musikwissenschaftlichen Seminar*, Humboldt-University Berlin, Germany

Publications

Music of Taiwanese *Luantan Xiao-Xi.*（台灣亂彈小戲音樂初探） Journal of Chinese Ritual, Theatre and Folklore (Taipei) **106**, 1-29, 1997. [in Chinese]

A comparison between Chinese *Nan-Xi* and French Opera Comique: a case of the structure of drama music *He-To* and vaudeville finale.（中國南戲與法國喜歌劇中的程式美典比較—以合頭與 vaudeville finale 的戲劇音樂結構為例） Arts Review **8**, (Taipei) 163-185, 1997. [in Chinese, with English summary]

Kunqu opera in folk-culture: a case of Taiwanese *Luantan* opera "*Qinqiong Daotongqi*".（庶民文化中的崑曲—台灣亂彈戲《秦瓊倒銅旗》）Shu Yenji Memorial Conference (Taipei), 1998. [in Chinese]

Relating Taiwanese *Luantan Fulu-Xi* to *Luantan* music system.（論台灣亂彈戲「福路」聲腔屬於「亂彈腔系」—唱腔與劇目的探討）Journal of Chinese Ritual, Theatre and Folklore (Taipei) **124**, 43-88, 2000. [in Chinese]

Physics of overtone-singing. Science Monthly (Taipei) **375**, 209-216, 2001. [in Chinese]

Relating the harmonic-rich sound of the Chinese flute (*dizi*) to the cubic nonlinearity of its membrane. Stockholm Music Acoustics Conference 2003, August 6–9, pp. 303-306.

Talks

"Intonation in folk music: Methodological considerations and practice", Graduate Institute of Musicology, National Taiwan University, 2000.

"Physics of overtone-singing", Department of Physics, University of Potsdam, 2003.

Eidestattliche Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt zu haben und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.