บทที่ 5

การทดลองและประเมินผล

การสังเคราะห์เสียงพูดด้วยวิธีแอลพีซี คุณภาพของเสียงที่ได้และจำนวนข้อมูลที่ใช้ในการสังเคราะห์เสียงขึ้นอยู่กับองค์ประกอบสำคัญคือ

1. ความละเอียดในการแปลงสัญญาณระหว่างข้อมูลเชิงเลขกับสัญญาณอนาลอก
2. ความถี่ในการสุ่มสัญญาณ
3. จำนวนออร์เดอร์ของแอลพีซีฟิลเตอร์
4. จำนวนแซมเปิลของสัญญาณในหนึ่งเฟรม

คุณภาพของเสียงกับจำนวนข้อมูลจะส่งผลตรงข้ามเสมอคือ การผลิตเสียงให้ได้คุณภาพดีต้องใช้ข้อมูลจำนวนมากและเมื่อข้อมูลมีจำนวนมากการนำไปใช้งานทำได้ไม่สะดวก โดยทั่วไปการสังเคราะห์เสียงจะกำหนดให้คุณภาพเสียงอยู่ในระดับใช้งานได้หรือสามารถสื่อความหมายจากการฟัง (Inteligible) โดยพยายามให้ข้อมูลเสียงมีจำนวนน้อยที่สุด

การเลือกองค์ประกอบ 4 ข้อข้างต้นอาศัยผลการทดลองจากเอกสารอ้างอิงเป็นแนวทางเพื่อวางข้อกำหนดในการทดลอง จากนั้นทำการทดลองเพื่อหาองค์ประกอบที่เหมาะสมสำหรับการผลิตเสียงพูดต่อๆ ไป การวางข้อกำหนดในการทดลองสรุปได้ดังนี้

1. การทดลองในงานวิจัยนี้ ใช้ความละเอียดในการแปลงสัญญาณเท่ากับ 12 บิท ทั้งในส่วนของการนำสัญญาณเสียงพูดมาวิเคราะห์และในส่วนของการผลิตเสียง ความละเอียดในการแปลงสัญญาณขนาดนี้มีขนาดของเสียงรบกวนที่เกิดจากการควอนไทซ์คิดเป็นค่า Signal to Noise Ratio เท่ากับ 72 db เทียบกับระบบพีซีเอม เสียงพูดที่มีคุณภาพดีสำหรับงานทั่วไปจะมีค่า SNR อยู่ประมาณ 60 db [4]

2. ความถี่ในการสุ่มสัญญาณ เลือกความถี่เท่ากับ 10 kHz จากเหตุผลที่ว่าความถี่ฟอร์แมนท์ที่สำคัญของเสียงพูดสำหรับเสียงที่มีคุณภาพจะอยู่ในย่านความถี่ไม่เกิน 5 kHz [8]

3. จำนวนออร์เดอร์ของแอลพีซีฟิลเตอร์ จำนวนออร์เดอร์ในที่นี้หมายความถึงจำนวนที่เหมาะสม (Optimum) เกิดจากการพิจารณาระหว่างคุณภาพของเสียงกับปริมาณข้อมูล ในเรื่องนี้จากเอกสารอ้างอิงพบว่า มีหลายความคิดเห็นเช่น

- B.S. Atal และ S.L. Hanauer[20] สรุปว่า จำนวนออร์เดอร์เท่ากับ 12 เหมาะสมที่สุด จากการทดลองการวิเคราะห์-สังเคราะห์เสียงด้วยจำนวนออร์เดอร์ระหว่าง 2 ถึง 18 พบว่าคุณภาพของเสียงพูดพิจารณาจากการฟังเมื่อจำนวนออร์เดอร์เท่ากับ 12 ขึ้นไป มีความแตกต่างกันน้อยมากและการสังเคราะห์เสียงด้วยจำนวนออร์เดอร์ต่ำกว่า 10 ทำให้เสียงนาสิกด้อยคุณภาพอย่างชัดเจน

- จากเอกสารอ้างอิง [13] กล่าวว่า จำนวนออร์เดอร์จะเท่ากับ 10 ถือว่าเหมาะสมที่สุดจากการเปรียบเทียบค่า Maximum Likelihood Ratio.

- จากเอกสารอ้างอิง [7,p.419] สรุปว่าจำนวนออร์เดอร์ของแอลพีซีฟิลเตอร์ที่เหมาะสมขึ้นกับย่านความถี่ของสัญญาณ จำนวนออร์เดอร์จะเท่ากับ 2 ต่อย่านความถี่ 1 kHz เช่น ย่านความถี่เท่ากับ 5 kHz หรือสุ่มสัญญาณด้วยความถี่ 10 kHz จำนวนออร์เดอร์เท่ากับ 10 ร่วมกับจำนวนออร์เดอร์อีก 3 ถึง 4 สำหรับจำลองคุณสมบัติของเส้นเสียงและการกระจายเสียงจากริมฝีปาก รวมเป็น 13 ถึง 14 ออร์เดอร์สำหรับสัญญาณที่มีย่านความถี่ 5 kHz การทดลองในงานวิจัยนี้กำหนดจำนวนออร์เดอร์อยู่ระหว่าง 2 ถึง 15 เพื่อให้ครอบคลุมแนวความคิดต่างๆ ที่กล่าวมา

4. จำนวนแซมเปิลในหนึ่งเฟรม จำนวนแซมเปิลในหนึ่งแฟรมขึ้นอยู่กับธรรมชาติของเสียงพูดในเรื่องของการเปลี่ยนแปลงคุณสมบัติตามเวลาและอัตราการสุ่มสัญญาณ เช่น ถ้าเสียงพูดมีคุณสมบัติค่อนข้างคงที่ไม่เปลี่ยนแปลงในช่วงเวลา 10 ms และความถี่ในการสุ่มสัญญาณเสียงเท่ากับ 10 kHz ดังนั้นเฟรมหนึ่งจะมีจำนวนแซมเปิลเท่ากับ 100 ในความเป็นจริงช่วงเวลาที่คุณสมบัติของเสียงเปลี่ยนแปลงแต่น้อยขึ้นอยู่กับภาษา สำเนียงการพูด ความเร็วในการพูด และอื่นๆ ดังนั้นในเรื่องนี้จำเป็นต้องอาศัยการทดลอง ในงานวิจัยนี้ใช้จำนวนแซมเปิลในหนึ่งเฟรมระหว่าง 100 ถึง 300 แซมเปิลหรือคิดเป็นช่วงเวลาจะเท่ากับ 10 ถึง 30 ms ต่อหนึ่งเฟรม

<u>วัตถุประสงค์ของการทดลอง</u>

คือต้องการผลิตเสียงพูดภาษาไทยเป็นคำๆ ให้มีคุณภาพใช้งานได้โดยพยายามให้ข้อมูลเสียงมีจำนวนน้อยที่สุด ดังนั้นแนวโน้มคือพยายามให้การสังเคราะห์เสียงใช้จำนวนออร์เดอร์น้อยที่สุดและให้จำนวนแซมเปิลต่อหนึ่งเฟรมมีจำนวนมากที่สุด

<u>ขั้นตอนการทดลอง</u> แบ่งเป็น 3 ขั้นตอนคือ

1. ทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ โดยจำนวนออร์เดอร์มีค่าต่างๆ ระหว่าง 2 ถึง 15 สรุปจำนวนออร์เดอร์ที่เหมาะสม

2. ทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ โดยจำนวนแซมเปิลในหนึ่งเฟรมมีค่าต่างๆ ระหว่าง 100 ถึง 300 สรุปจำนวนแซมเปิลในหนึ่งเฟรมที่เหมาะสม

3. ทดลองสังเคราะห์เสียงเป็นคำๆ ของตัวเลขหนึ่งถึงสิบ

<u>การเปรียบเทียบคุณภาพของเสียง</u> กระทำได้ 2 วิธีคือ

1. ใช้ค่าผิดพลาดนอร์มัลไลซ์ (Normalized Error) ในแต่ละเฟรม ค่าผิดพลาดนอร์มัลไลซ์ คืออัตราส่วนระหว่างพลังงานของสัญญาณค่าผิดพลาดที่ได้จากอินเวิร์สฟิลเตอร์กับพลังงานของสัญญาณเสียงจริงในหนึ่งเฟรม หรือ

$$\text{Normalized Error} = \frac{\sum_{n=0}^{N-1} e^2(n)}{\sum_{n=0}^{N-1} s^2(n)} \qquad (5.1)$$

ซึ่งในการวิเคราะห์ด้วยวิธีพาร์คอร์ ค่าผิดพลาดนอร์มัลไลซ์จะเท่ากับ

$$\text{NE parcor} = \prod_{i=1}^{M} (1-k_i^2) \qquad (5.2)$$

เมื่อ M คือจำนวนออร์เดอร์ของฟิลเตอร์ [7,p.426]

ค่าผิดพลาดนอร์มัลไลซ์จะเป็นตัวบอกถึงผลของการทำนายจากตัวทำนายเชิงเส้นว่า ภายใต้เงื่อนไขค่าผิดพลาดยกกำลังสองรวมในหนึ่งเฟรมน้อยที่สุดจะมีค่าผิดพลาดเท่าใด ค่าผิดพลาดน้อยแสดงว่าการทำนายทำได้ดี และในการสังเคราะห์ก็จะมีแนวโน้มได้เสียงที่มีคุณภาพดี

2. การเปรียบเทียบโดยการฟัง เป็นการเปรียบเทียบที่ดีที่สุด เพราะผลลัพธ์ที่ต้องการคือเสียงซึ่งสามารถใช้งานได้ ฟังแล้วสื่อความหมายโดยไม่จำเป็นที่สัญญาณเสียงจะต้องเหมือนกับสัญญาณเสียงต้นแบบ

5.1 การทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ด้วยจำนวนออร์เดอร์ต่าง ๆ

การทดลองแบ่งเป็น 2 ส่วนคือ

1. วิเคราะห์เสียงสระ อา ซึ่งเป็นเสียงสระที่มีคุณสมบัติค่อนข้างคงที่ไม่ค่อยเปลี่ยนแปลงตามเวลา การทดลองเริ่มจากคำนวณหาสัมประสิทธิ์พาร์คอร์ด้วยโปรแกรม LPCX ทั้งหมด 7 ครั้ง โดยกำหนดจำนวนออร์เดอร์ของฟิลเตอร์เท่ากับ 2,4,6,8,10,12 และ 15 ตามลำดับ การวิเคราะห์ทั้ง 7 ครั้ง กำหนดจำนวนแซมเปิลในหนึ่งเฟรม (Analysis Frame Lenght) เท่ากับ 160 แซมเปิล จำนวนแซมเปิลที่ผ่าน Window เท่ากับ 200 แล้วนำสัญญาณเสียงสระ อา มาคำนวณในโปรแกรม SIFTX เพื่อคำนวณหาคาบ จากนั้นนำข้อมูลพารามิเตอร์ของฟิลเตอร์ 7 ชุดร่วมกับข้อมูลคาบของสัญญาณมาสังเคราะห์เป็นสัญญาณเสียงด้วยโปรแกรม SYNTX จะได้สัญญาณเสียงสังเคราะห์ทั้งหมด 7 ไฟล์ นำสัญญาณมาแสดงเป็นรูปกราฟเมื่อเปรียบเทียบรูปร่างของสัญญาณ และนำสัญญาณเสียงที่สังเคราะห์ได้ไปคำนวณ FFT เพื่อเปรียบเทียบสเปกตรัมของสัญญาณ แล้วนำค่าสัมประสิทธิ์พาร์คอร์มาคำนวณหาค่าผิดพลาดนอร์มัลไลซ์แสดงเป็นกราฟเปรียบเทียบระหว่างจำนวนออร์เดอร์ต่าง ๆ

2. ทำการวิเคราะห์เสียงพูดของคำว่า กา ทั้งคำด้วยโปรแกรม LPCX และโปรแกรม SIFTX แล้วทำการสังเคราะห์เสียงด้วยโปรแกรม SYNTX ในลักษณะเดียวกับ ข้อ 1. เปรียบเทียบเสียงที่ได้ด้วยการฟังโดยอาศัยโปรแกรม Signal Editor นำข้อมูลที่ใช้ในการสังเคราะห์เสียงไปสังเคราะห์เสียงด้วยภาคประมวลผลสัญญาณเปรียบเทียบเสียงที่ได้ด้วยการฟัง

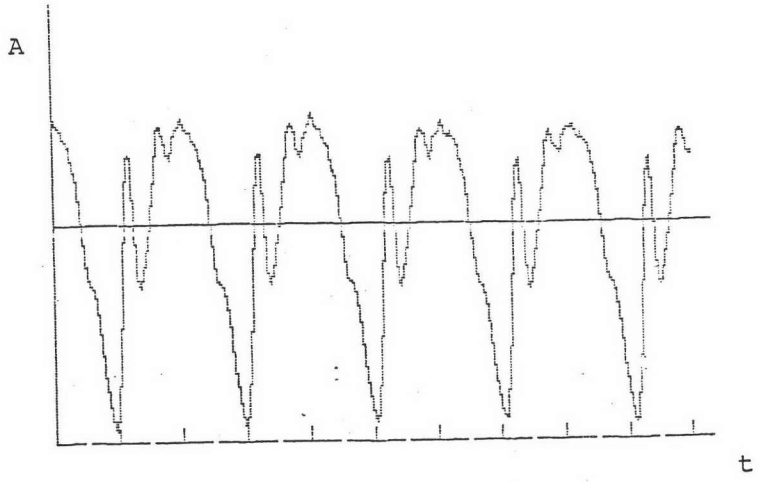
ผลการทดลองแสดงให้เห็นในรูป 5.1 ถึง รูป 5.9 ในรูป 5.1 แสดงรูปร่างของสัญญาณ สเปกตรัมของสัญญาณ และสเปกตรัมที่ผ่านการทำให้ราบเรียบ (Smooth) ของสัญญาณเสียง สระอา ซึ่งเป็นเสียงต้นแบบ รูป 5.2 ถึง รูป 5.8 แสดงรูปร่างของสัญญาณ สเปกตรัมของสัญญาณและสเปกตรัมที่ผ่านการทำให้ราบเรียบของเสียงสังเคราะห์ โดยกำหนดจำนวนออร์เดอร์ของฟิลเตอร์ต่างๆ

จากรูป 5.1 ก. สัญญาณของเสียงพูดสระอา มีคาบเท่ากับ 56 แซมเปิล จะเห็นอิทธิพลของความถี่ฟอร์แมนท์ที่หนึ่งได้ชัดเจน สังเกตจากยอด 3-4 ยอดในหนึ่งคาบ รูป 5.1 ข แสดงสเปกตรัมของเสียงสระอา จากรูปจะเห็นยอดเล็กๆ จำนวนมากมายซึ่งคือส่วนประกอบความถี่ของต้นกำเนิดเสียงหรือเสียงจากเส้นเสียง (Vocal Cords) ยอดแรกของสเปกตรัม คือ ยอดของความถี่หลักมูลซึ่งมีค่าประมาณ 180 Hz ยอดเล็กๆ ถัดมาจะตรงกับความถี่ฮาร์มอนิกส์ที่ 1,2,3,... ซึ่งค่อยๆ ลดอิทธิพลลงตามความถี่ รูป 5.1 ค คือ สเปกตรัมในรูป ข ที่ถูกทำให้ราบเรียบได้เป็น Spectral Envelope ของสเปกตรัมในรูป ข หรือเป็นผลตอบความถี่ของทางเดินเสียง

จากรูปจะเห็นยอดของฟอร์แมนท์ที่ 1,2,3 และ 4 ชัดเจน ฟอร์แมนท์ที่ 1 มีความถี่ประมาณ 700 Hz ฟอร์แมนท์ที่ 2 มีความถี่ประมาณ 1300 Hz เว้นระยะไปจนถึงบริเวณ 2900 Hz จะเป็นความถี่ฟอร์แมนท์ที่ 3 และฟอร์แมนท์ที่ 4 มีความถี่ประมาณ 3700 Hz ใช้รูป 5.1 เป็นหลักแล้วเปรียบเทียบกับรูป 5.2 ถึง 5.8 ซึ่งเป็นรูปของสัญญาณเสียงสังเคราะห์ เปรียบเทียบรูป ก ด้วยกันจะพบว่าเมื่อจำนวนออร์เดอร์เพิ่มขึ้นส่วนประกอบทางความถี่สูงจะชัดขึ้นเรื่อยๆ เมื่อเปรียบเทียบรูป ข ด้วยกัน จะเห็นว่าในส่วนของยอดเล็กๆ ซึ่งเป็นความถี่ฮาร์มอนิกส์ต่างๆ ของต้นกำเนิดเสียงในที่นี้ คือ Pulse Train ความถี่เท่ากับคาบของเสียง จะมีลักษณะคล้ายคลึงกันหมด การลดทอนขนาดของฮาร์มอนิกส์เป็นไปอย่างช้าๆ ทำให้เสียงแตกพร่า พิจารณารูป ค ของรูป 5.1 ถึง รูป 5.8 ซึ่งเป็นสเปกตรัมที่ผ่านการทำให้ราบเรียบ รูป 5.2 ค จำนวนออร์เดอร์เท่ากับ 2 จะมียอดอยู่ยอดเดียว ซึ่งไม่สามารถแยกได้ว่าเป็นความถี่ฟอร์แมนท์ใด ในลักษณะนี้เสียงจะมี คุณภาพต่ำมากจนฟังไม่รู้เรื่อง รูป 5.3 ค เริ่มเห็นยอดความถี่ฟอร์แมนท์บริเวณ 800 Hz ซึ่งเป็นผลจากความถี่ฟอร์แมนท์ที่ 1 และ 2 รวมกัน รูป 5.4 ค จำนวนออร์เดอร์เท่ากับ 6 ยอดแรกจะเห็นชัดเจนขึ้นแต่ก็ยังแยกไม่ได้ระหว่างความถี่ฟอร์แมนท์ที่ 1 และที่ 2 นอกจากนั้นยังมียอดของฟอร์แมนท์ที่ 3 และที่ 4 อยู่รวมกันอีกหนึ่งยอดที่บริเวณ 3 kHz รูป 5.5 ค จำนวนออร์เดอร์เท่ากับ 8 สเปกตรัมยังคงมีลักษณะคล้ายคลึงกับรูป 5.4 ค เมื่อจำนวนออร์เดอร์เพิ่มขึ้นเป็น 10 จากรูป 5.6 ค จะเห็นยอดของฟอร์แมนท์ครบทั้ง 4 ยอด รูป 5.7 ค และรูป 5.8 ค มีลักษณะคล้ายคลึงกัน ยอดของฟอร์แมนท์ที่ 1 และที่ 2 เห็นชัดเจนขึ้น เปรียบเทียบรูป 5.8 ค กับรูป 5.1 ค ลักษณะยอดฟอร์แมนท์ต่างๆ คล้ายคลึงกันมากและคุณภาพเสียงที่ได้จะดีที่สุด

ผลการคำนวณหาค่าผิดพลาดนอร์มัลไลซ์แสดงไว้ในรูป 5.9 ซึ่งในการคำนวณอาศัยค่าเฉลี่ยของเฟรมที่ติดกัน 5 เฟรม ค่าผิดพลาดนอร์มัลไลซ์ลดลงเมื่อจำนวนออร์เดอร์เพิ่มขึ้นจาก 2 ถึง 15 ซึ่งจะลดลงอย่างรวดเร็วในช่วงจำนวนออร์เดอร์เท่ากับ 2 ถึง 6 และจะมีค่าใกล้เคียงกันระหว่างจำนวนออร์เดอร์เท่ากับ 6 กับ 8 ค่าผิดพลาดนอร์มัลไลซ์ลดลงมากเมื่อจำนวนออร์เดอร์เท่ากับ 10 และจะค่อยๆ ลดลงเมื่อจำนวนออร์เดอร์เพิ่มขึ้นเป็น 12 และ 15

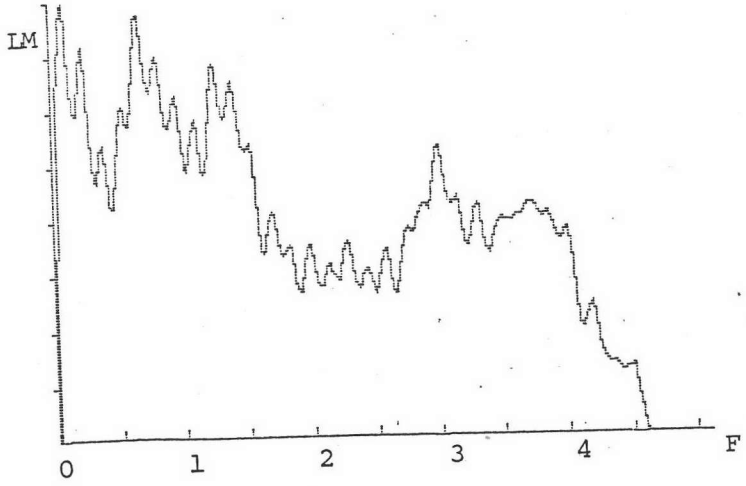
จากการทดลองฟังเสียงที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 2 ถึง 6 เสียงยังมีคุณภาพไม่ดี จำนวนออร์เดอร์เท่ากับ 8 เสียงมีคุณภาพพอใช้ จำนวนออร์เดอร์เท่ากับ 10 คุณภาพเสียงดีขึ้นมาก จำนวนออร์เดอร์เท่ากับ 12 ถึง 15 เสียงมีคุณภาพใกล้เคียงกัน จากการทดลองสรุปได้ว่า จำนวนออร์เดอร์เท่ากับ 10 เหมาะสมที่สุด



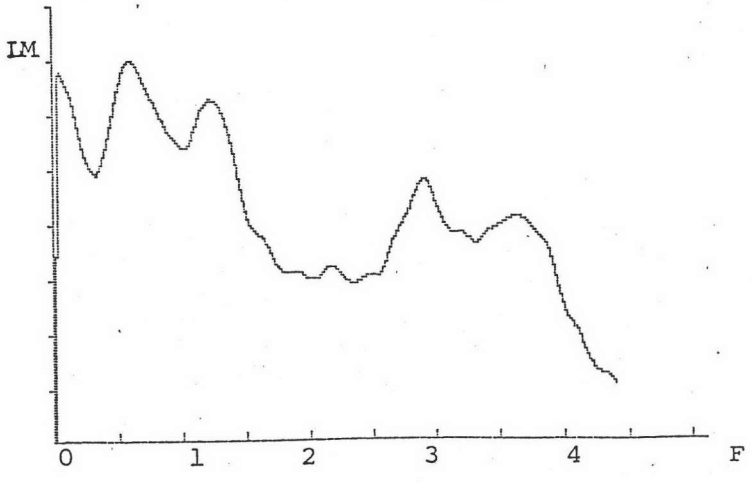
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

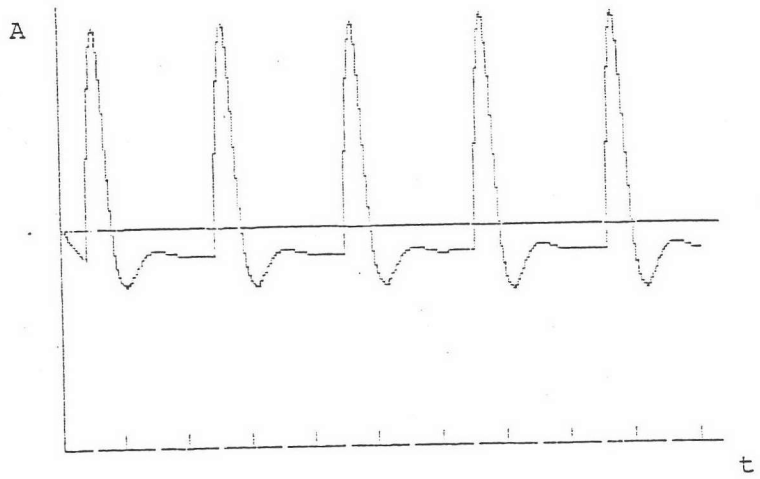


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

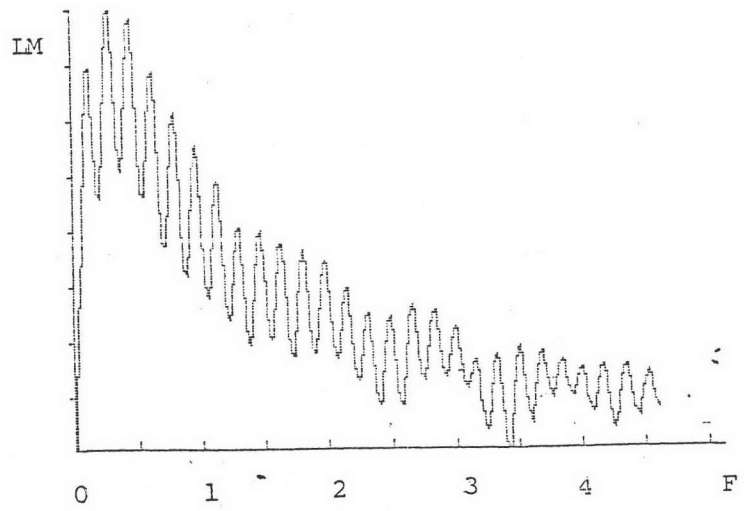
รูป 5.1 สัญญาณและสเปกตรัมของเสียงสระอาต้นแบบ



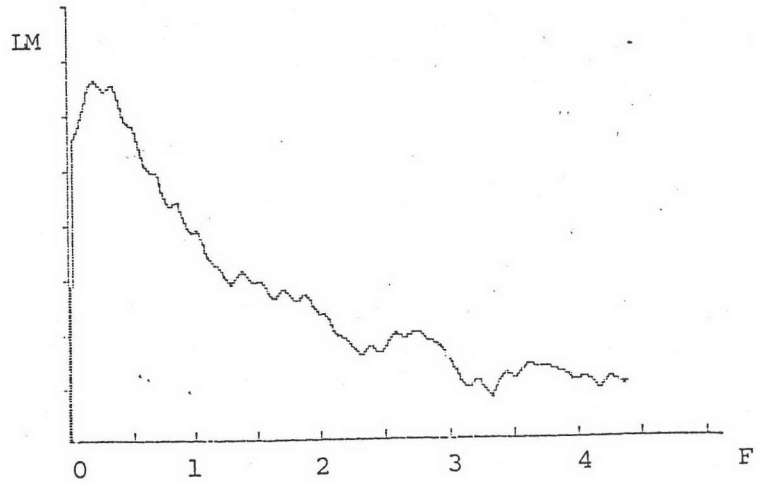
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

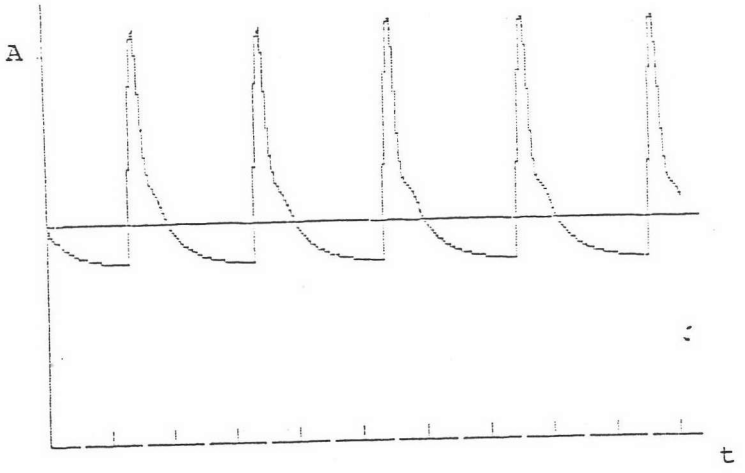


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

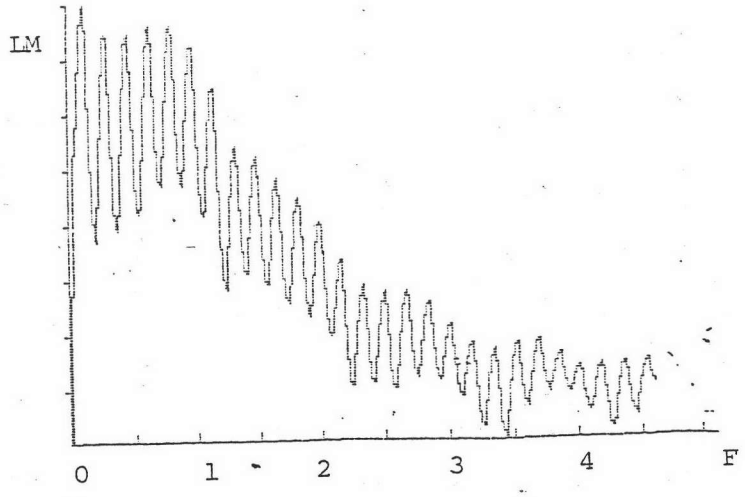
รูป 5.2 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 2



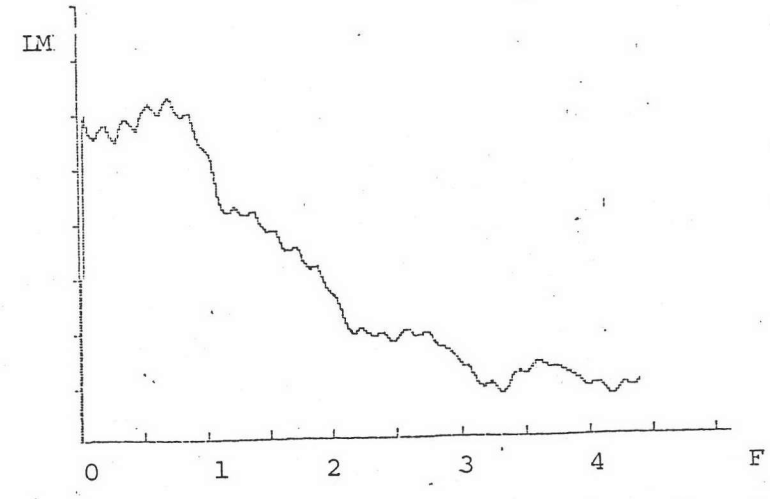
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

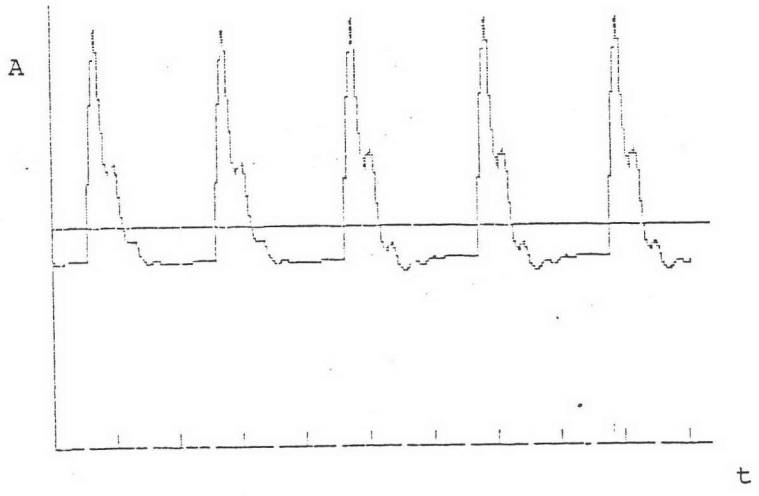


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

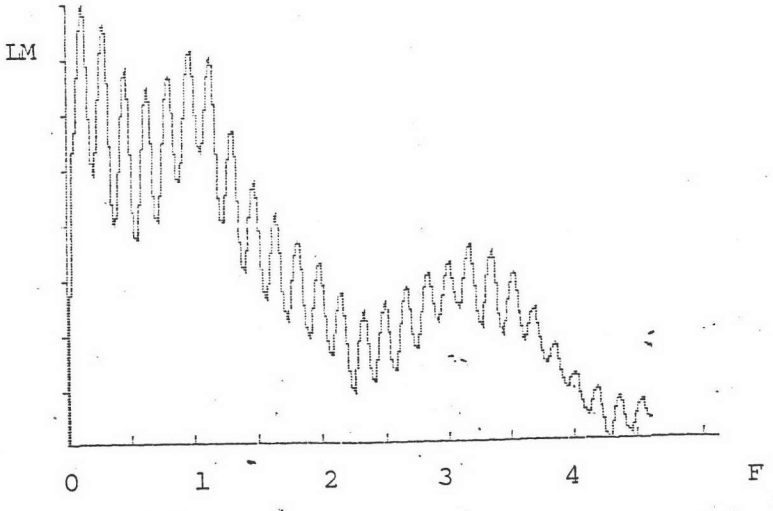
รูป 5.3 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 4



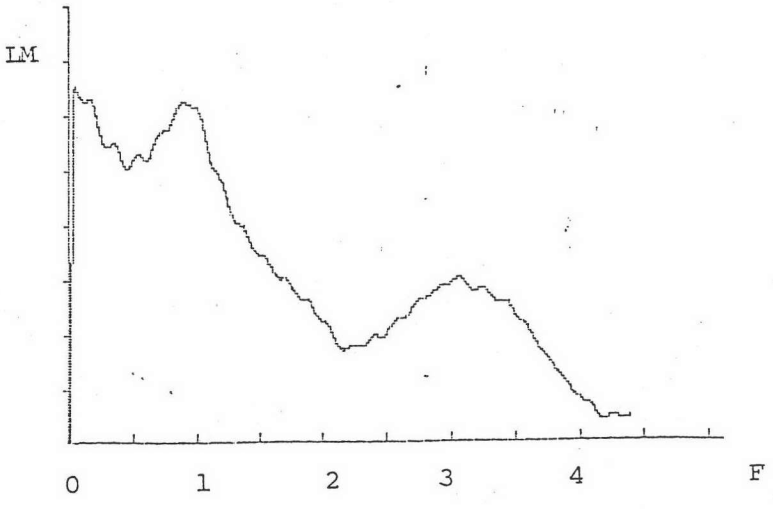
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

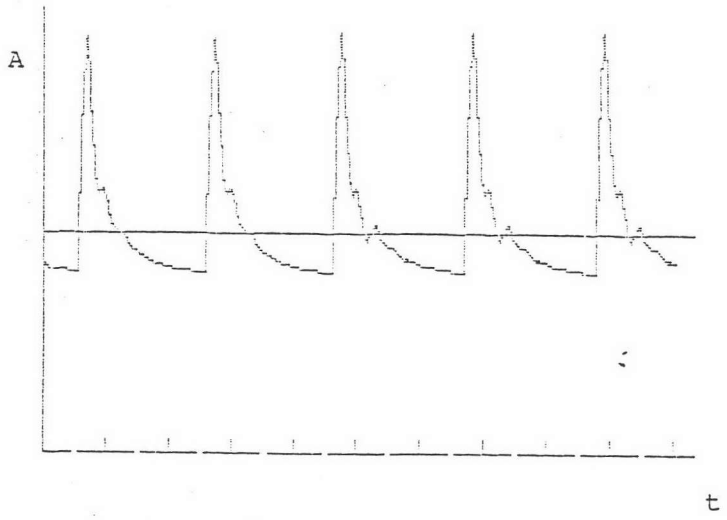


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

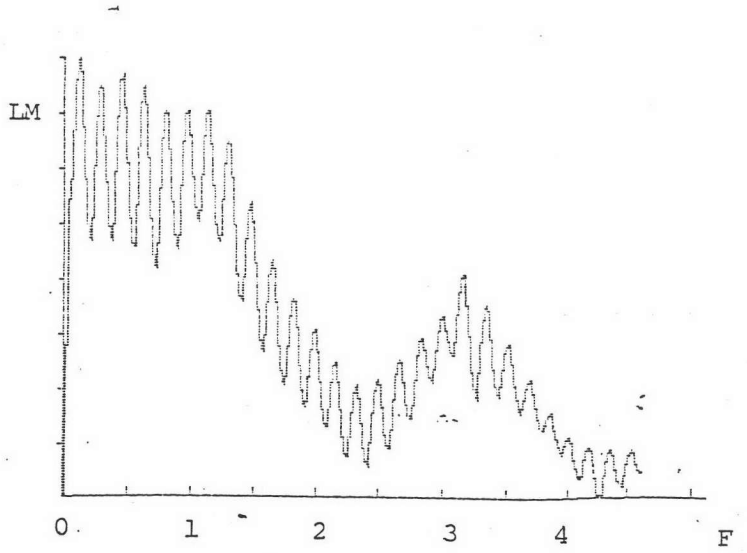
รูป 5.4 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 6



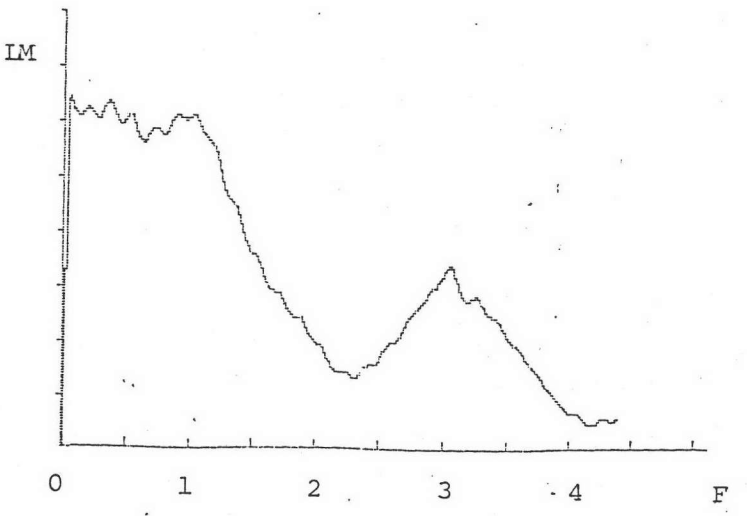
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

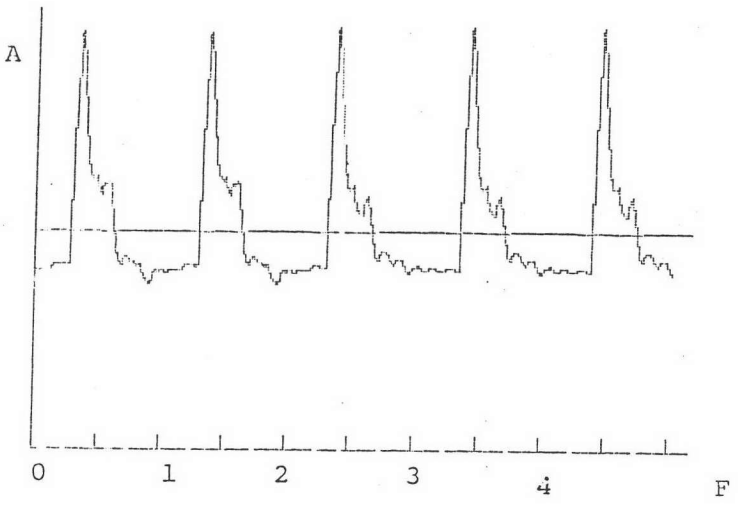


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

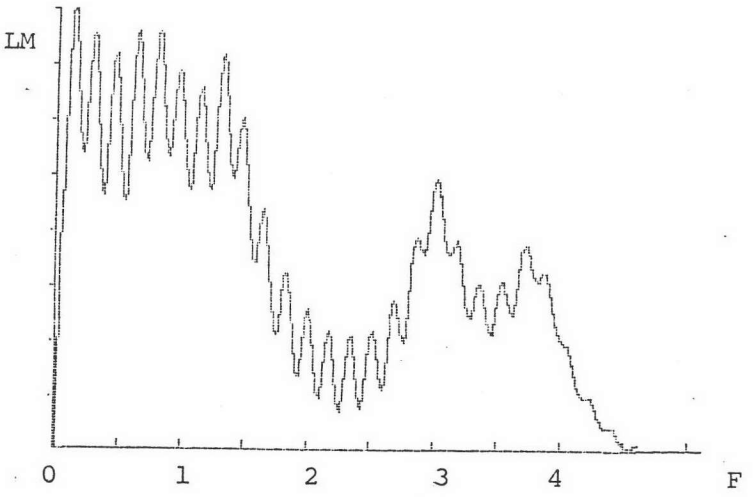
รูป 5.5 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 8



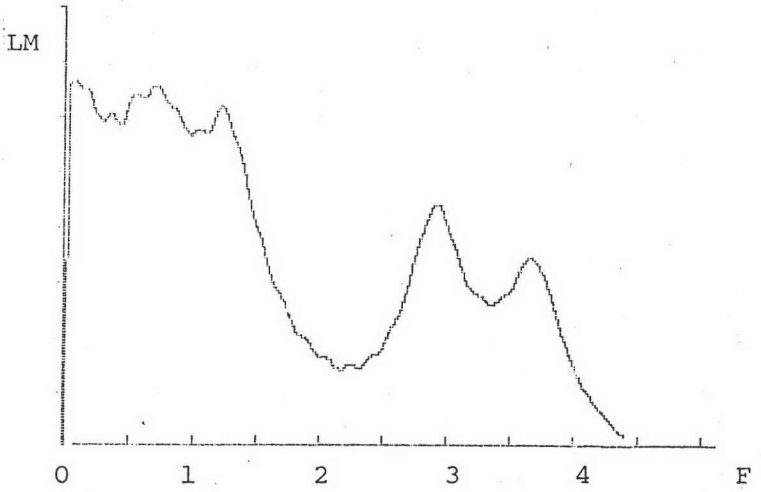
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

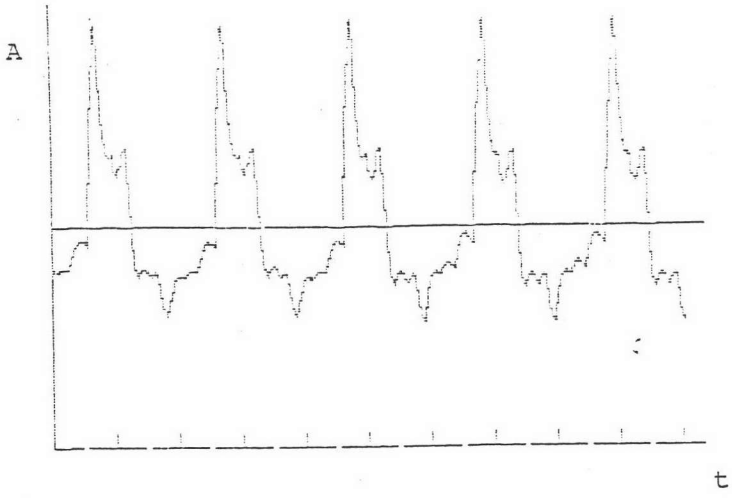


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

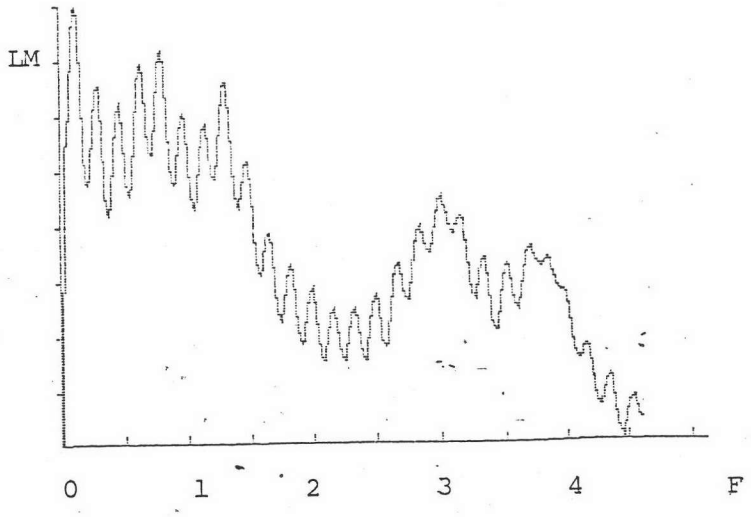
รูป 5.6 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 10



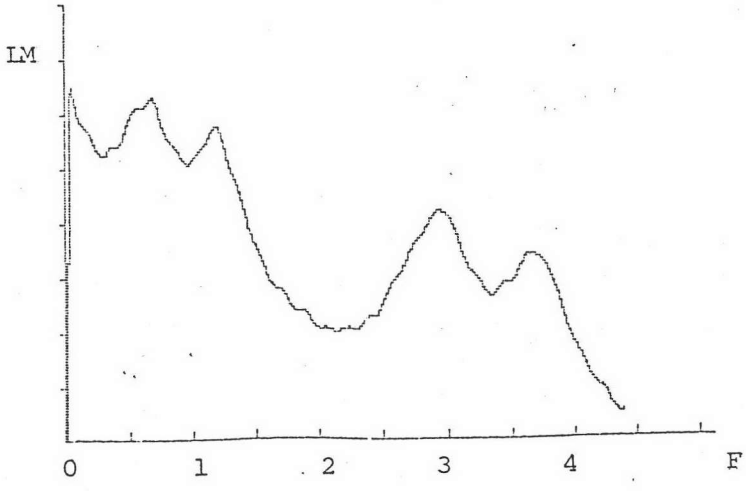
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

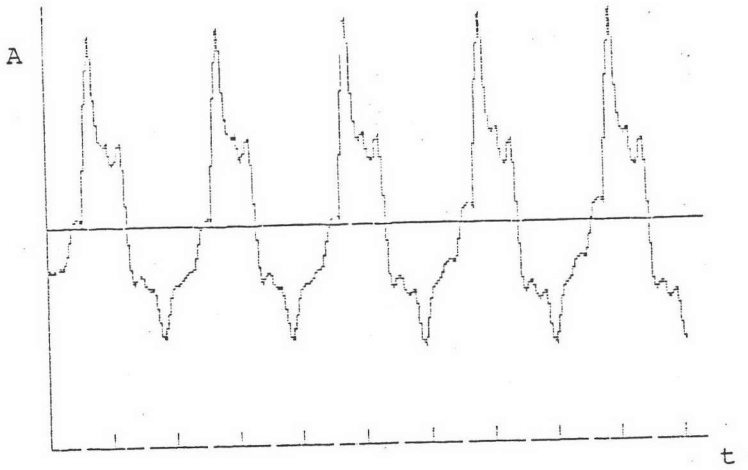


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

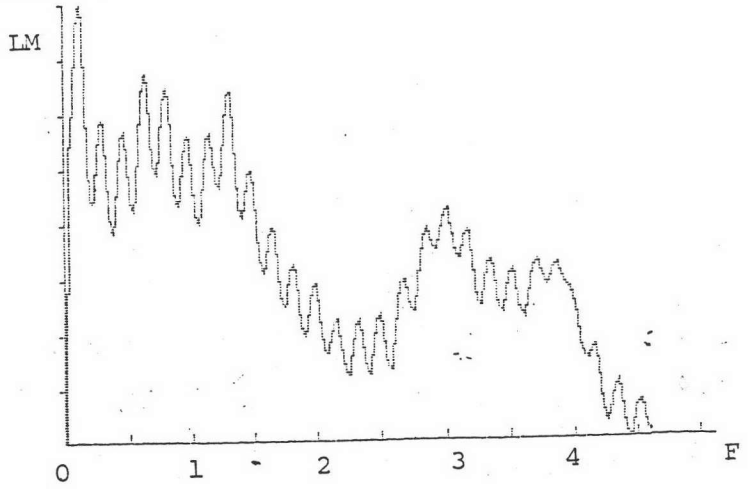
รูป 5.7 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 12



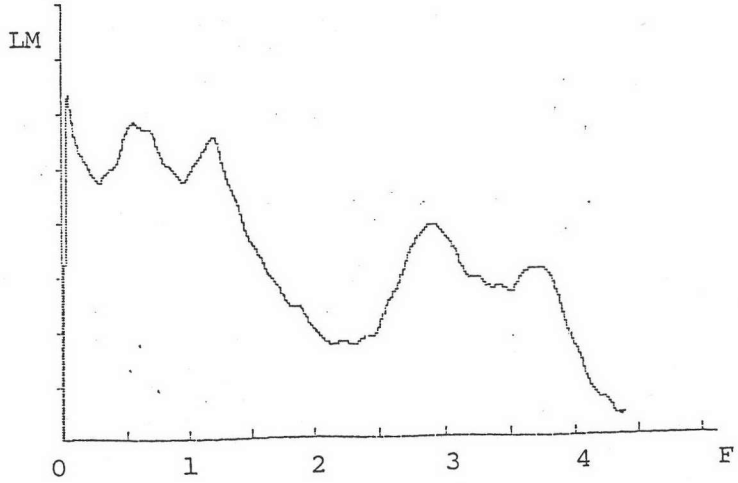
ก) สัญญาณเสียง

A = Amplitude

t = time 2.7 ms/DIV.



ข) สเปกตรัม

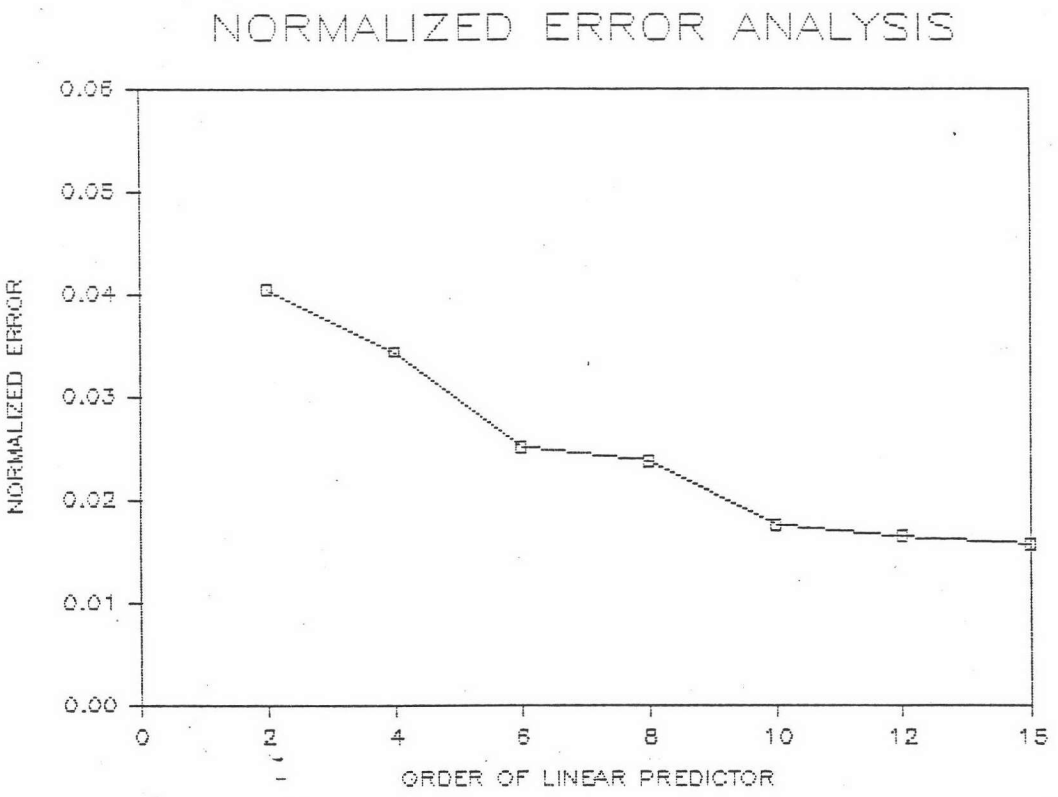


ค) สเปกตรัมที่ผ่านการทำให้ราบเรียบ

LM = Log - Magnitude

F = Frequency (kHz)

รูป 5.8 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ จำนวนออร์เดอร์เท่ากับ 15



Experiment	Filter Order	Normalized Error
1	2	0.04047
2	4	0.0344
3	6	0.02513
4	8	0.02383
5	10	0.0176
6	12	0.0165
7	15	0.01577

รูป 5.9 ผลการทดลองคำนวณค่าผิดพลาดนอร์มัลไลซ์ของเสียงสระอา

5.2 การทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ ด้วยจำนวนแซมเปิลในหนึ่งเฟรมต่าง ๆ

การทดลองกระทำโดยนำเสียงพูดคำว่า "ก้า" มาวิเคราะห์ทั้งคำด้วยโปรแกรม LPCX และโปรแกรม SIFTX จำนวน 5 ครั้ง แต่ละครั้งกำหนดจำนวนแซมเปิลในหนึ่งเฟรมหรือ N เท่ากับ 100, 150, 200, 250 และ 300 ตามลำดับ การทดลองทั้ง 5 ครั้ง กำหนดจำนวนออร์เดอร์เท่ากับ 10 นำข้อมูลที่ได้ไปสังเคราะห์เสียงด้วยโปรแกรม SYNTX ทดสอบคุณภาพเสียงด้วยการฟังโดยอาศัยโปรแกรม SIGNAL EDITOR และนำข้อมูลที่ใช้ในการสังเคราะห์เสียงไปสังเคราะห์เสียงด้วยภาคประมวลผลสัญญาณเปรียบเทียบเสียงที่ได้ด้วยการฟัง

เสียงพูดคำว่า "ก้า" จะแตกต่างกับคำว่า "กา" เนื่องจากมีการเปลี่ยนแปลงของคาบหรือความถี่หลักมูลของเสียง ดังนั้นในช่วงเสียงสระ คุณสมบัติของรูปคลื่นจะเปลี่ยนแปลงตามเวลา รูปที่ 5.10 แสดงรูปร่างสัญญาณเสียงพูดคำว่า "ก้า" กับค่าของคาบหรือความถี่หลักมูลเสียง ซึ่งเปลี่ยนแปลงตามลักษณะของวรรณยุกต์โท ถ้าจำนวนแซมเปิลในหนึ่งเฟรมมีจำนวนมาก จะเกิดการเปลี่ยนแปลงคุณสมบัติของสัญญาณภายในเฟรม ทำให้การทำนายมีค่าผิดพลาดสูงขึ้นได้

ในส่วนของการวิเคราะห์ค่าผิดพลาดนอร์มัลไลซ์ แบ่งเสียงพูดทั้งคำเป็น 5 ส่วนย่อยๆ จากต้นคำถึงท้ายคำในจำนวนเฟรมที่เท่าๆ กัน คำนวณหาค่าผิดพลาดนอร์มัลไลซ์ที่เฟรมแรกของแต่ละส่วน ผลลัพธ์ที่ได้แสดงใน รูป 5.11 ถึง รูป 5.15 จากผลการทดลองจะเห็นว่าค่าของ N อยู่ระหว่าง 100 ถึง 200 ช่วงต้นของคำ ค่าผิดพลาดนอร์มัลไลซ์มีค่าต่ำและค่อยๆ เพิ่มขึ้นตามส่วนของคำ เมื่อค่าของ N เท่ากับ 250 และ 300 ค่าผิดพลาดนอร์มัลไลซ์จะมีค่ามากตั้งแต่ต้นคำ ที่เป็นเช่นนี้ตั้งข้อสังเกตว่าเฟรมมีขนาดใหญ่ ในช่วงต้นของคำถ้าการเริ่มต้นของเฟรมอยู่ในตำแหน่งที่ไม่เหมาะสมจะทำให้มีค่าผิดพลาดสูง เมื่อนำค่าผิดพลาดนอร์มัลไลซ์ของทุกส่วนมาหาค่าเฉลี่ยจะได้ผลดังรูป 5.16 จากรูป 5.16 จะเห็นว่าค่าผิดพลาดนอร์มัลไลซ์เพิ่มขึ้นเมื่อค่าของ N เพิ่มขึ้นจาก 150 ถึง 300 ซึ่งก็สอดคล้องกับเหตุผลที่ว่า จำนวนแซมเปิลในหนึ่งเฟรมมีจำนวนมาก คุณสมบัติของเสียงจะมีการเปลี่ยนแปลงภายในเฟรม ส่วนกรณีที่ค่าผิดพลาดนอร์มัลไลซ์ เมื่อ N เท่ากับ 100 มีค่ามากกว่าเมื่อ N เท่ากับ 150 เนื่องจากเมื่อค่าของ N มีขนาดใกล้เคียงกับคาบของเสียงจะทำให้ค่าผิดพลาดมีค่ามาก [7]

จากการทดลองฟังเสียงที่ได้จากการสังเคราะห์ พบว่าค่าของ N อยู่ระหว่าง 100 ถึง 250 คุณภาพเสียงใกล้เคียงกัน กรณี N เท่ากับ 300 คุณภาพเสียงลดลงอย่างเห็นได้ชัด สรุปค่าของ N อยู่ระหว่าง 200 ถึง 250 ให้ผลการสังเคราะห์เสียงที่น่าพอใจ และเมื่อเปรียบเทียบจำนวนข้อมูลเสียงที่ได้จากรูป 5.16 การลดทอนจะอยู่ประมาณ 15-19 เท่า การลดทอนจำนวนข้อมูล

(Data Reduction) คำนวณจากจำนวนแซมเปิลของเสียงกับจำนวนข้อมูลที่ใช้สังเคราะห์เสียงหรือ

$$\text{ค่าลดทอนข้อมูล} = \frac{\text{จำนวนแซมเปิลสัญญาณเสียงในหนึ่งเฟรม}}{\text{จำนวนข้อมูลที่ใช้ในการสังเคราะห์เสียงในหนึ่งเฟรม}}$$

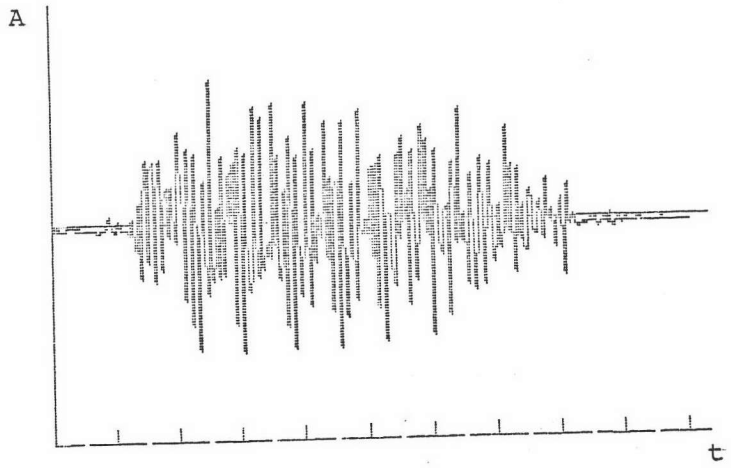
ตารางที่ 5.1 แสดงค่าลดทอนข้อมูลเมื่อกำหนดจำนวนออร์เดอร์ของฟิลเตอร์ M ระหว่าง 8 ถึง 15 และจำนวนแซมเปิลในหนึ่งเฟรม N มีค่าระหว่าง 100 ถึง 300

M \ N	100	150	200	250	300
8	9.1	13.6	18.2	22.7	27.3
10	7.7	11.5	15.4	19.2	23.1
12	6.7	10.0	13.3	16.7	20.0
15	5.6	8.3	11.1	13.9	16.7

ตารางที่ 5.1 ค่าลดทอนข้อมูลเมื่อกำหนดค่า M และ N ต่าง ๆ

ในการทดลอง จำนวนข้อมูลเสียงในหนึ่งเฟรมจะเท่ากับ 13 จากการกำหนดจำนวนออร์เดอร์เท่ากับ 10 ค่าลดทอนข้อมูลเมื่อเลือก N เท่ากับ 200 และ 250 มีค่าเท่ากับ 15.4 และ 19.2 ตามลำดับ ซึ่งนับว่าอยู่ในระดับที่ดี

สรุป จำนวนแซมเปิลในหนึ่งเฟรมอยู่ระหว่าง 200 ถึง 250 เป็นจำนวนที่เหมาะสม พิจารณาทั้งด้านคุณภาพของเสียงและจำนวนข้อมูลที่ใช้ในการสังเคราะห์

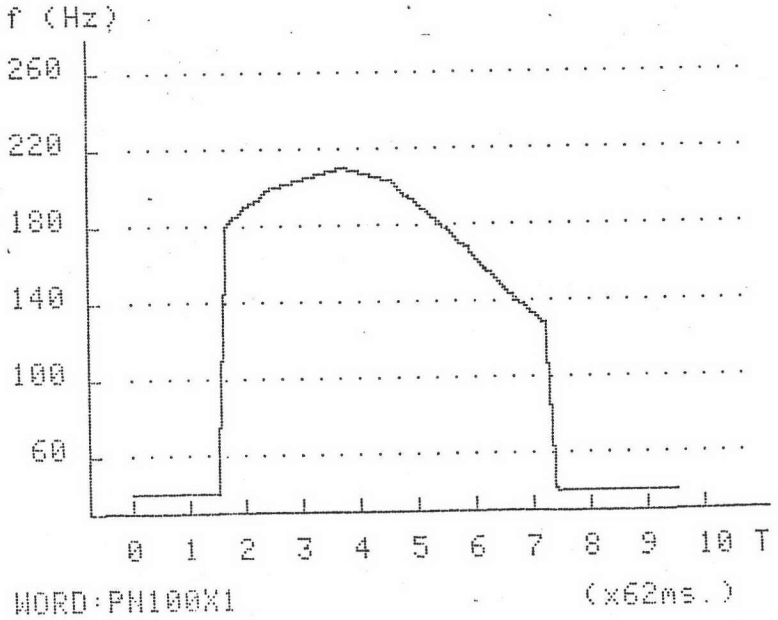


ก) สัญญาณเสียง

A = Amplitude

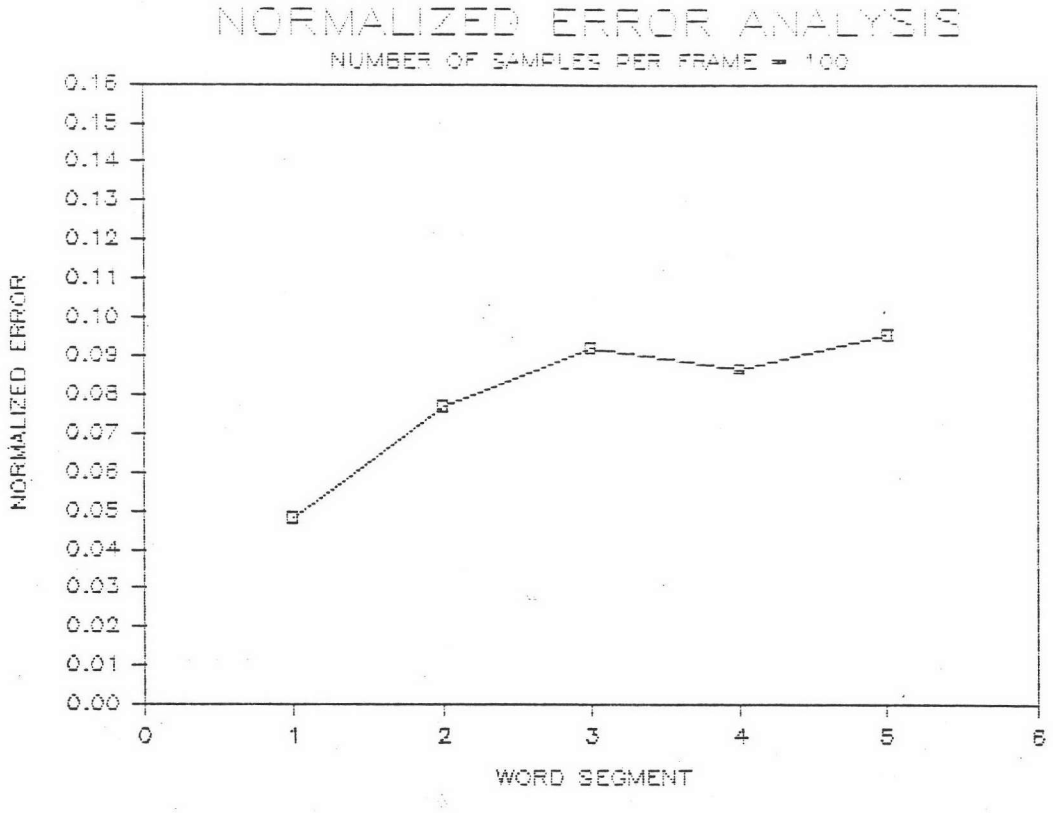
t = time

40 ms/DIV.



ข) ความถี่หลักมูล

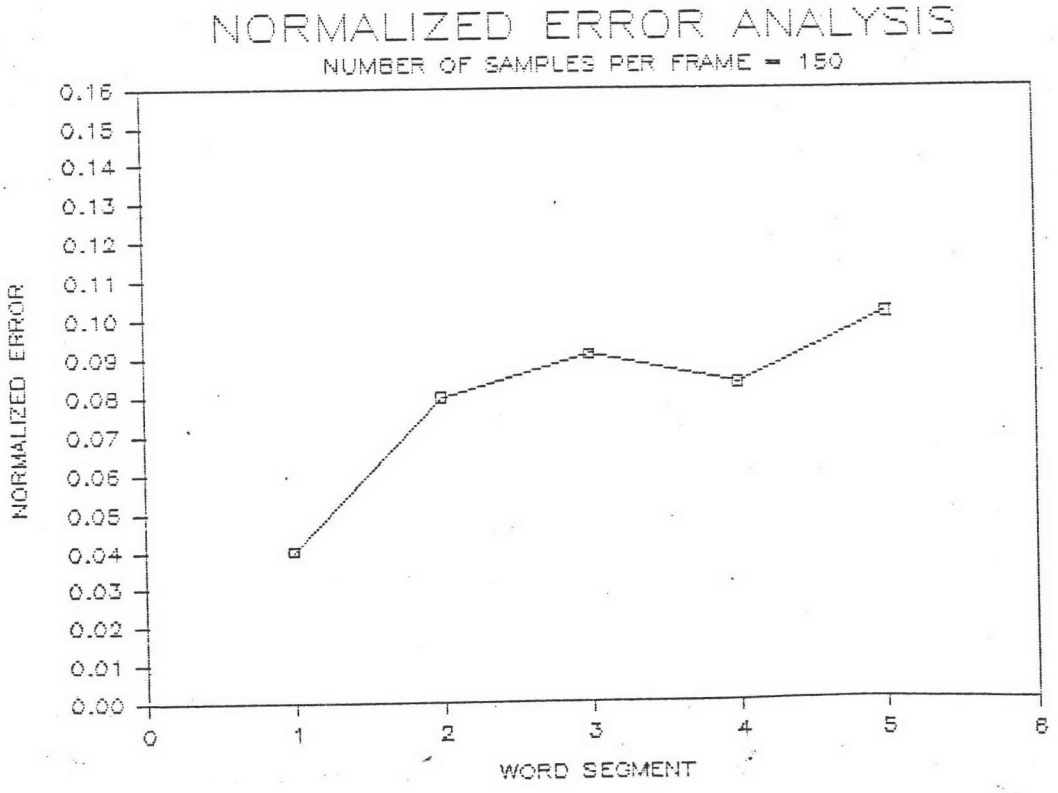
รูป 5.10 สัญญาณและความถี่หลักมูลของเสียงพูดคำว่า"ก้า"



N = 100

SEGMENT	NORMALIZED ERROR
1	0.048147
2	0.079646
3	0.092121
4	0.086775
5	0.095712

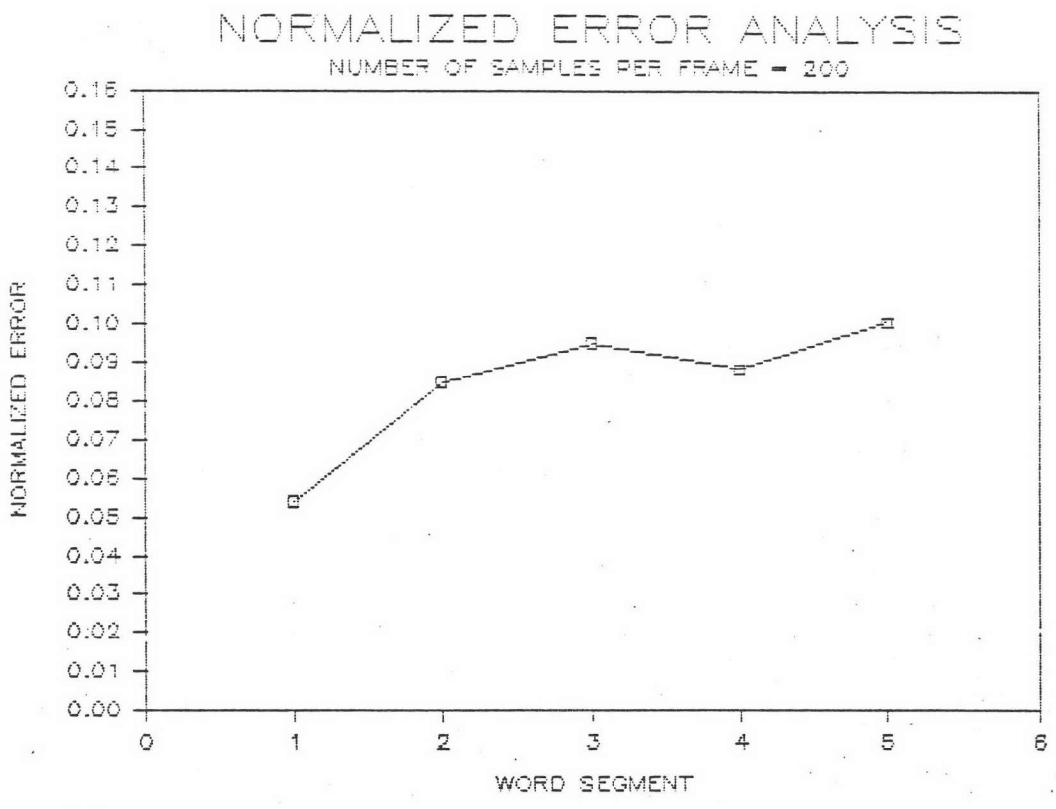
รูป 5.11 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า"
เมื่อจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 100



N = 150

SEGMENT	NORMALIZED ERROR
1	0.040037
2	0.079954
3	0.091130
4	0.083566
5	0.101560

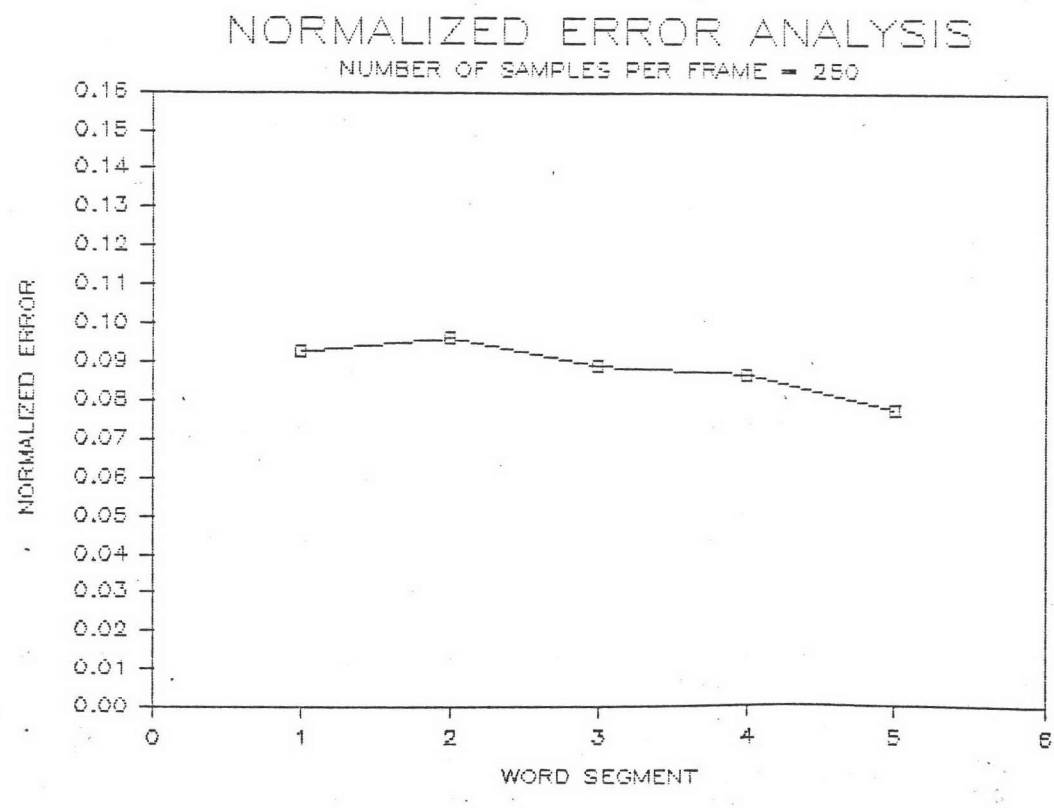
รูป 5.12 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า"
เมื่อจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 150



N = 200

SEGMENT	NORMALIZED ERROR
1	0.053822
2	0.084945
3	0.094975
4	0.088237
5	0.100482

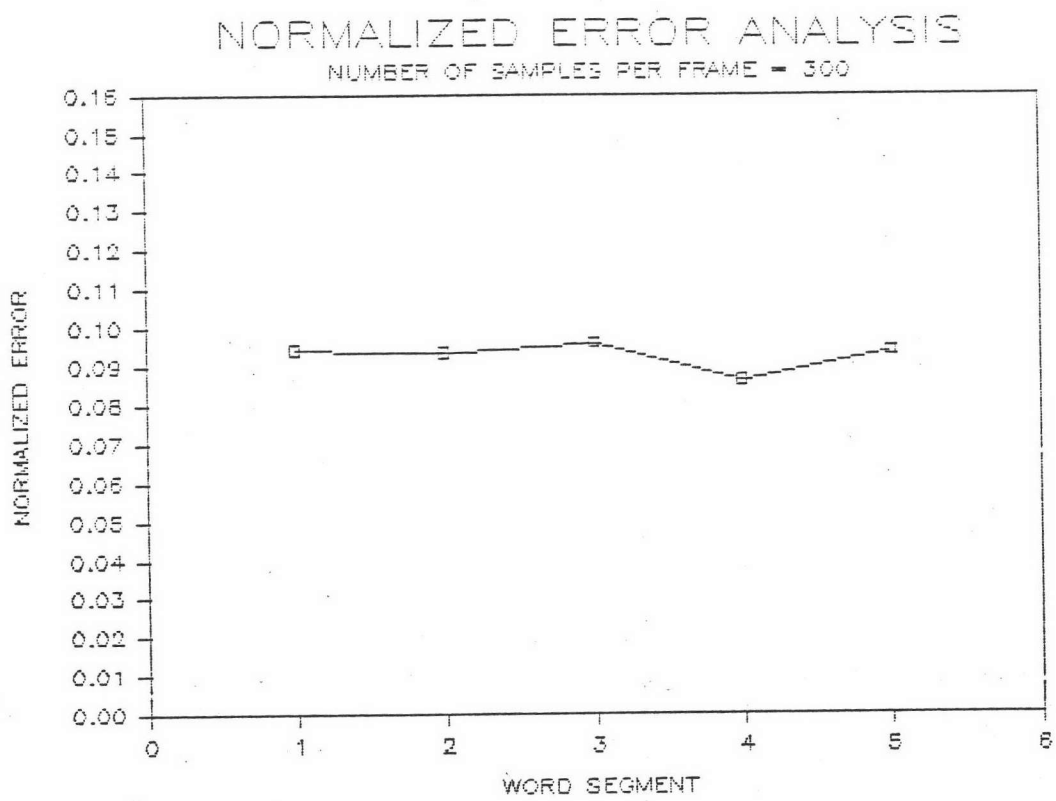
รูป 5.13 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า"
เมื่อจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 200



N = 250

SEGMENT	NORMALIZED ERROR
1	0.092666
2	0.096314
3	0.089150
4	0.089182
5	0.077910

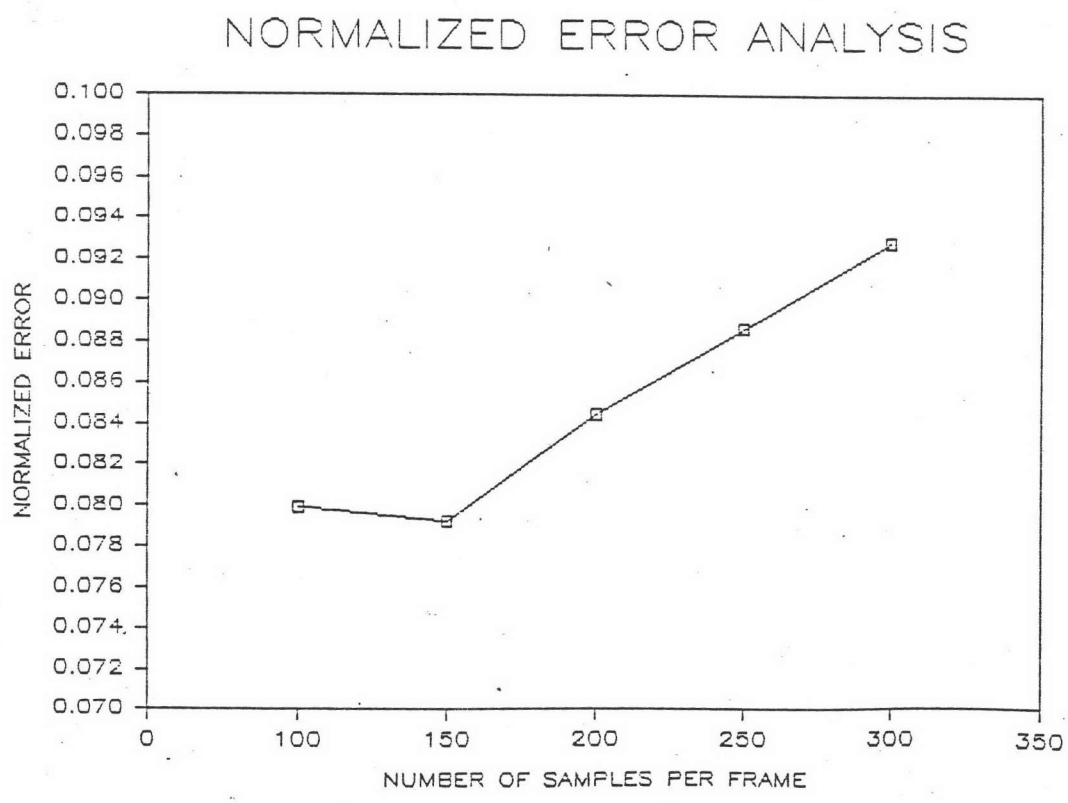
รูป 5.14 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า"
เมื่อจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 250



N = 300

SEGMENT	NORMALIZED ERROR
1	0.094283
2	0.093575
3	0.096090
4	0.086415
5	0.093919

รูป 5.15 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า"
เมื่อจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 300



N	DATA REDUCTION	AV. NORMALIZED ERROR
100	7.7	0.079944
150	11.5	0.079250
200	15.4	0.084492
250	19.2	0.088644
300	23.1	0.092854

M = 10; Number of speech data in 1 frame = 13

รูป 5.16 เปรียบเทียบค่าเฉลี่ยค่าผิดพลาดนอร์มัลไลซ์จากทุกส่วนของคำพูด"ก้า" รวมทั้งค่าลดทอนข้อมูล

5.3 การทดลองวิเคราะห์และสังเคราะห์เสียงพูดของตัวเลขหนึ่งถึงสิบ

การทดลองนี้เริ่มจากสุ่มสัญญาณเสียงพูดของคำว่า "หนึ่ง" ถึง "สิบ" โดยควบคุมการพูดให้สม่ำเสมอ นำเสียงพูดที่ได้ผ่านการคำนวณด้วยโปรแกรม LPCX และโปรแกรม SIFTX โดยกำหนดจำนวนออร์เดอร์ของฟิลเตอร์เท่ากับ 10 และจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 200 ทุกคำ นำข้อมูลที่ได้ผ่านการแปลงข้อมูลด้วยโปรแกรม DCONX และทดสอบการสังเคราะห์เสียงจากภาคประมวลผลสัญญาณ ผลลัพธ์ที่ได้สรุปไว้ในตารางในรูป 5.17

WORD	LENGTH (ms)	TOTAL FRAME	DATA (WORD)
1	400	19	251
2	560	28	368
3	560	28	368
4	400	19	251
5	360	18	238
6	220	10	134
7	220	10	134
8	230	15	199
9	400	19	251
10	220	10	134

M = 10, N = 200

รูป 5.17 ผลการทดลองวิเคราะห์-สังเคราะห์คำพูด"หนึ่ง" ถึง "สิบ"