Effect of F0 fluctuation and amplitude modulation of natural vowels on vowel identification in noisy environments
Kentaro Ishizuka and Kiyoaki Aikawa
NTT Communication Science Labs., JAPAN
{ishizuka@atom,aik@idea}.brl.ntt.co.jp

Abstract
This paper describes findings showing that the fundamental frequency (F0) fluctuation and amplitude modulation (AM) included in natural vowels contribute to improving the vowel identification rate in the presence of interferer sounds. A vowel identification experiment revealed that even very small F0 fluctuations and AM of vowels significantly improved the vowel identification rate in the presence of interferer sounds. In addition, it was found that the effect is not the result of reducing the masked threshold. A vowel detection experiment revealed that the effect produced by reducing the threshold depends on the kinds of interferers, and that the vowel identification rates do not correlate with the masked thresholds.

Background

• Findings on speech perception in noisy environments
  – The frequency modulation (FM) of a vowel increases the prominence of that vowel in the presence of other vowels [e.g. McAdams, J. Acoust. Soc. Am. 86, 2148-2159 (1989)].
  – The amplitude modulation (AM) of vowels contributes to improving speech perception [e.g. Carrel and Opie, Percept. and Psychophys. 52, 437-445 (1992)].

• The sustained part of a single natural vowel includes;
  – FM: rapid, narrow-band, 1/f characteristic F0 fluctuation (jitter)
  – AM: uncomodulated amplitude fluctuation across the frequencies

• There have been few findings related to the effect of the F0 fluctuation and AM of natural vowels in noisy environments.
Motivation

Can the jitter of natural vowels improve the vowel identification rate?

– Rapid, narrow-band, $1/f$ characteristic F0 fluctuation of target vowels improves the vowel identification rate in the presence of a harmonic complex tone as an interferer sound even at very low SNRs [Ishizuka and Aikawa, *J. Acoust. Soc. Am.* 110, 2680 (2001)].

– The result suggests that the jitter of natural vowels can improve the vowel identification rate in such a periodic, harmonic noise in the real world.

Can the AM of natural vowels (*uncomodulated across frequency*) improve the vowel identification rate?

– The effect of comodulated amplitude fluctuation across the frequencies has been investigated in the field of auditory grouping [e.g. Hall and Grose, *J. Acoust. Soc. Am.* 84, 1669-1675 (1988)].

– In contrast, the effect of uncomodulated amplitude fluctuation across the frequencies (such as AM of natural vowels) has not been considered.

Purpose

An examination of whether the F0 fluctuation and AM of natural vowels can improve the vowel identification rate in the presence of interferer sounds.

Method

A vowel identification experiment was conducted to measure the vowel identification rate as a function of male vowel synthesis conditions corresponding to whether the vowels have the jitter and/or the AM of natural vowels in the presence of interferer sounds.
Vowel Identification Experiment

Method
– Vowel identification rates in the presence of interferer sounds were measured as a function of male vowel synthesis conditions.
– Vowel synthesis conditions (four conditions):
  • no modulation (without jitter and AM)
  • with only the AM of a natural vowel
  • with only the jitter of a natural vowel
  • with both the jitter and AM of a natural vowel
– Vowels were synthesized by LPC filtering the glottal pulse sequences.
– Vowels (five kinds): Japanese vowels /a/, /i/, /u/, /e/, and /o/
– Interferer sounds (two kinds): a harmonic complex tone and a band noise

F0 Fluctuation and AM of natural vowels (used in this experiment)

F0 fluctuation (jitter)
Waveform of sustained part of natural vowel /a/

|------|------|------|------|------|------|------|------
| 7.79ms | 7.77ms | 7.81ms | 7.73ms | 7.85ms | 7.77ms | 7.90ms |...

Glottal pulse period length (ms)

F0 fluctuation (jitter) of natural vowel /a/
Mean F0: 127.7 Hz,
Jitter bandwidth: 5.4% of the mean F0

AM
Natural vowel
Constant-bandwidth filterbank
Half-wave rectified
Lowpass filtering
Cutoff frequency: 50 Hz

Center frequency: $N \cdot 127.7$ Hz,
Bandwidth: 127.7 Hz,
$N$: integer

AM in each frequency band of natural vowel
Interferer Sounds

Harmonic complex tone
- \( F_0 = 127.7 \text{ Hz} \) (= mean \( F_0 \) of vowels)
- 42 components
  - (5361.7 Hz at highest)
- Schroeder-negative phase
- SNR: 0, -3, and -6 dB

Band noise
- Cutoff frequency: 5361.7 Hz
- SNR: -6.5, -9.5, and –12.5 dB

Procedure
- Subjects were forced to choose one of the five Japanese vowels /a/, /i/, /u/, /e/, and /o/ as the stimulus sound.
- An experiment for each subject consisted of four experimental sessions corresponding to the four vowel synthesis conditions.
- The order of the sessions and the stimulus presentation in each session were randomized.
- The vowels were presented at 60 dB SPL.

Subjects
- 20 male and 20 female Japanese between 18 and 23 years old.
Vowel Identification Rate

In presence of harmonic complex tone

- The F0 fluctuation of the vowels increased the vowel identification rate.
- The AM of the vowels also increased the identification rate only when the vowel had F0 fluctuation.

In presence of band noise

- The F0 fluctuation of the vowels increased the vowel identification rate.
- The AM of the vowels did not increase the identification rate.

Masked Threshold Measurement

Purpose
- To confirm whether the increase in the vowel identification rates was due to the decrease in the masked threshold.

Method and Procedure
- The masked thresholds of vowel detection with the interferer sounds were measured as a function of the vowel synthesis conditions.
- The synthesized vowel /a/ was used as a signal vowel.
- The harmonic complex tone was presented at 60 dB SPL and the band noise was presented at 66.5 dB SPL (corresponding to the SNRs in the preceding identification experiment).
- 3-down-1-up 2AFC method.

Subjects
- 10 male and 10 female Japanese between 19 and 22 years old.
Masked Threshold

In presence of harmonic complex tone (60 dB SPL)

- The increase in the identification rate caused by the F0 fluctuation might be due to the decrease in the masked threshold.
- The increase in the identification rate caused by the AM was not due to the decrease in the masked threshold.

In presence of band noise (66.5 dB SPL)

- The increase in the identification rate was not due to the decrease in the masked threshold.

Conclusion and Discussion

- The F0 fluctuation of the vowels significantly increased the vowel identification rate in the presence of interferer sounds.

In the harmonic complex tone

- The increase in the identification rate might be due to the decrease in the threshold.
- The reason for the decrease in the masked threshold must be investigated – perhaps it occurred because of such factors as grouping by beating or harmonic cancellation.

In the band noise

- The increase in the identification rate was not due to the decrease in the threshold.
- The reason for this must be investigated.

- The AM of the vowels increased the vowel identification rate.

In the harmonic complex tone only when the vowel had F0 fluctuation

- The increase in the identification rate was not due to the decrease in the threshold, and must be due to something other than masking release such as CMR.