Text-to-Speech Tendency of Accent Errors in Japanese Verbs

Outline

1. Introduction/Research Purpose
2. Study Plan
3. Results and Discussion
4. Additional Experiment
5. Conclusion and Future Study Plan
6. References
7. Question and Answer

Text-to-Speech around us

TTS; Text-to-Speech (speech synthesis)

- There are a lot of TTS around us.
- Voice quality has improved in the last few years.

Figure1: Siri [1]

Figure2: Google Home [2]

Figure3: Hatsune Miku [3]

[1] iPhone Media. “Siri no uminooya ga kataru [genzai no Siri ni kaketeiru mono]” (n Japanese). https://iphone-mania.jp/news-205564/
[2] Robosuta. “Sayonara Google home・Googme mini?” (in Japanese). https://robotstart.info/2020/05/28/google-home-no-longer-available.html
[3] piaro.net. “Character” (in Japanese). https://piapro.net/pages/character

Let’s listening example speech of TTS (Siri)

❌ **However, prosodic mistakes were pointed out [4].**

Prosodic feature

Accents, Intonation, Speech speed, Voice tone, etc…

●Japanese judge word meaning **by the accent [5]**.

➡ **We should solve Accent mistake for smooth communication.**

[4] Masayuki Suzuki, et al (2013). "Jokentukikakurituba wo motita Nihongo-Tokyohogen no akusentoketugo-jidosuitei" (in Japanese). The IEICE Transactions. D, 96(3), 644-654

[5] Tadashi Sakamoto, et al (2017). "Nihongo-kyoiku heno michishirube dai 2 kan kotoba no shikumi wo shiru" (in Japanese). Bonjinsya

Research Purpose

Previous studies examined…

●Accent position estimating

●Natural language processing

●Waveform connection..

➡ Few studies searched what kind of accent error tendency exists in TTS around us.

Pursose of this study

To find the accent error tendency in TTS

Outline

1. Introduction/Research Purpose
2. **Study Plan**
3. Results and Discussion
4. Additional Experiment
5. Conclusion and Future Study Plan
6. References
7. Question and Answer

What is accent

- Judged by dropping position of pitch curve (F0) [6].
- Dropping position is called **accent nucleus.**

*Note: “ ` ” stands for nucleus.

- <u>e.g. *Tabe`ru (Eat)*</u>

Nucleus!

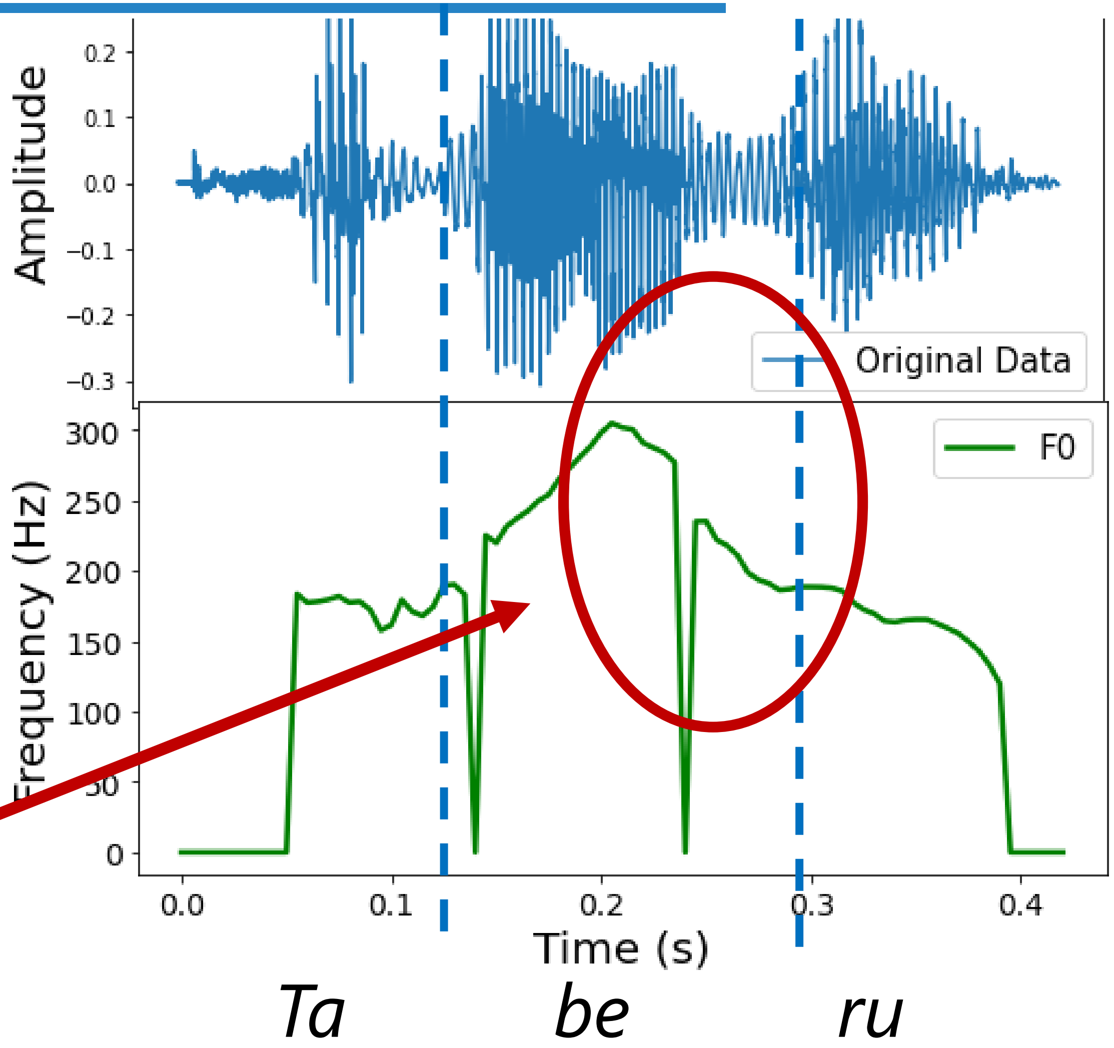


Figure4 (re): Wave form and F0 curve of “Taberu”

[6] Kyoko Takeuchi, et al (2019). "Tanosi Onseigaku" (in Japanese). Kuroshiosyuppan

Experiment Condition

- Focused on **3 mora (beat) verbs.** ← 3 mora means ○○○ in Japanese.
- 3 mora verbs has **2 accent types [6].**

1. Unaccented

- No nucleus
- e.g. *Kariru* (borrow)

2. Medial accented

- Have nucleus at the 2nd mora
- e.g. *Tabe`ru* (eat)

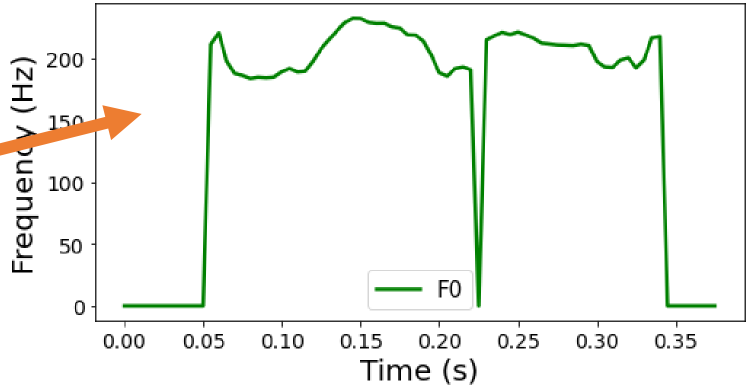


Figure5: F0 curve of “Kariru”

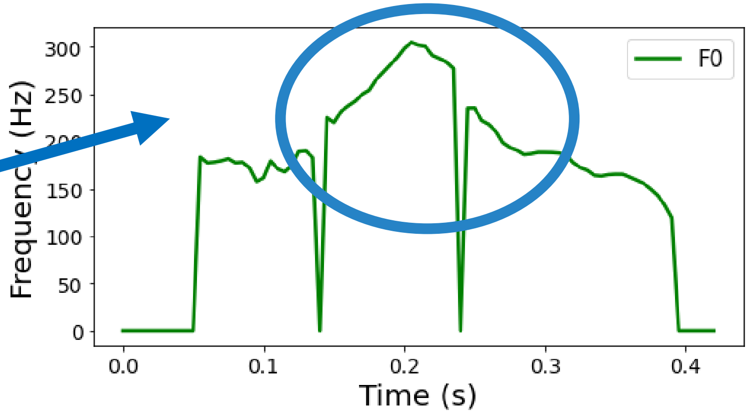


Figure6: F0 curve of “Taberu”

[6] Kyoko Takeuchi, et al (2019). "*Tanosi Onseigaku*" (in Japanese). Kuroshiosyuppan

The rules of accent variation

Table 1: The rules of accent variation [7]

	Imperfect	Continuative	Dictionary	Attributive
Unaccented	**○○○-nai.**	**○○○-ma`su.**	**○○○.**	**○○○-hito.**
Medial	**○○○-`nai.**	**○○○-ma`su.**	**○○`○.**	**○○`○-tito.**

	Conditional	Imperative	Volitional
Unaccented	**○○○-`ba.**	**○○○.**	**○○○-masyo`u.**
Medial	**○○○-`ba.**	**○○`○.**	**○○○-masyo`u.**

Ex. *Tabe`ru* + `nai → *Tabe`nai* (Imperfect form)

[7] NHK Broadcasting Culture Research Institute (2016). "NHK nihongo hatsuon akusent sinjiten" (in Japanese). NHK syuppan

Experiment Condition

Microsoft

&

aws

→ Readout!

60 unaccented | 60 medial

×

7 conjugated forms

- OJAD [10] was used to search target verbs.
- Reference F0 curve analyzed by Python.

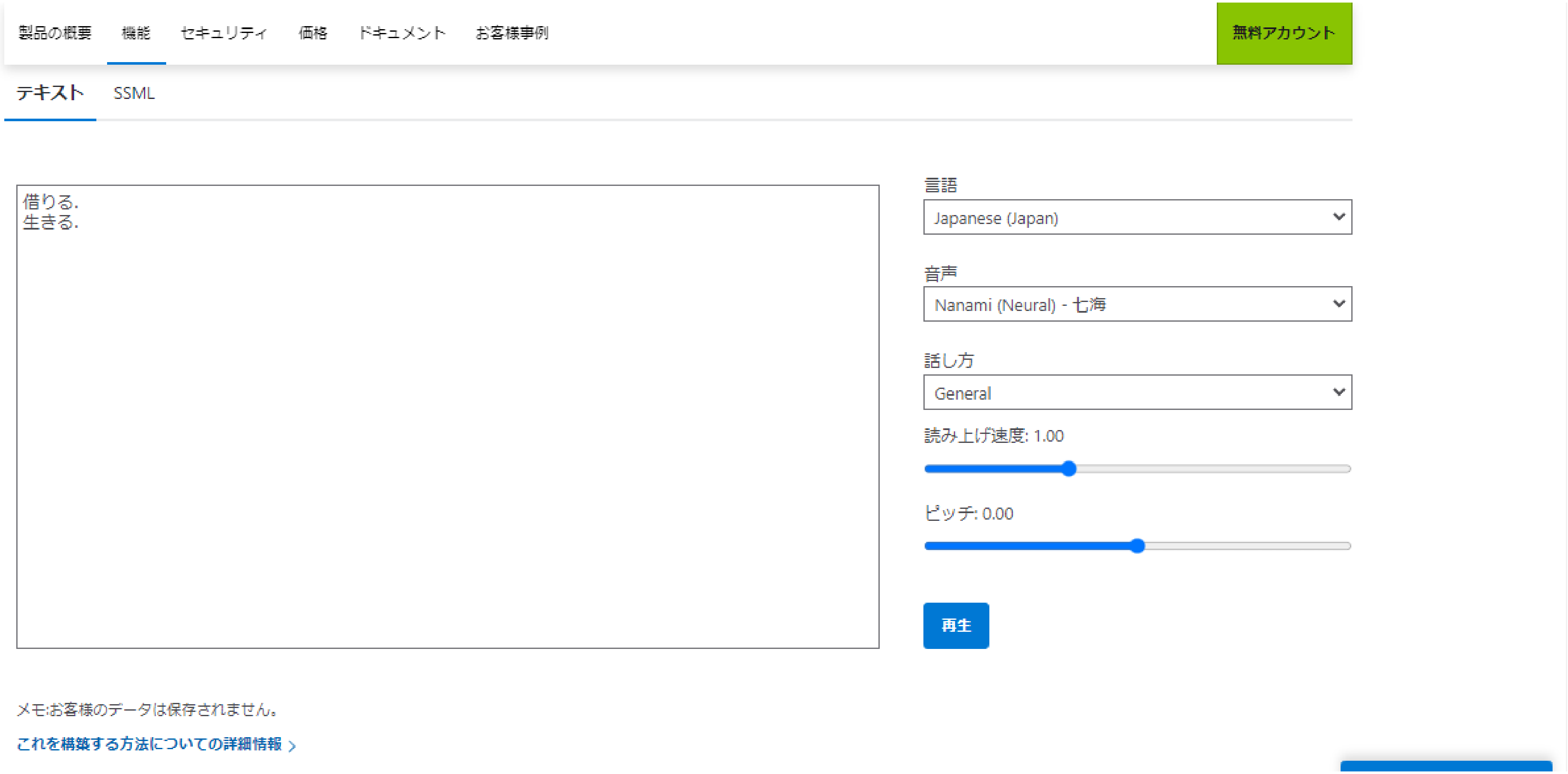
Examine the error tendency under each condition.

[8] Microsoft. "Text to Speech". https://azure.microsoft.com/ja-jp/services/cognitive-services/text-to-speech/
[9] Amazon. "Amazon Polly". https://cloud.google.com/text-to-speech?hl=ja
[10] Nobuaki Minematsu, et al (2013). "Online Japanese Accent Dictionary". https://www.gavo.t.u-tokyo.ac.jp/ojad/

Experiment Condition

We can try TTS online

Microsoft



借りる.
生きる.

言語: Japanese (Japan)

音声: Nanami (Neural) - 七海

話し方: General

読み上げ速度: 1.00

ピッチ: 0.00

メモ:お客様のデータは保存されません。

これを構築する方法についての詳細情報 >

[8] Microsoft. "Text to Speech". https://azure.microsoft.com/ja-jp/services/cognitive-services/text-to-speech/

Outline

1. Introduction/ Research Purpose
2. Study Plan
3. **Results and Discussion**
4. Additional Experiment
5. Conclusion and Future Study Plan
6. References
7. Question and Answer

Result

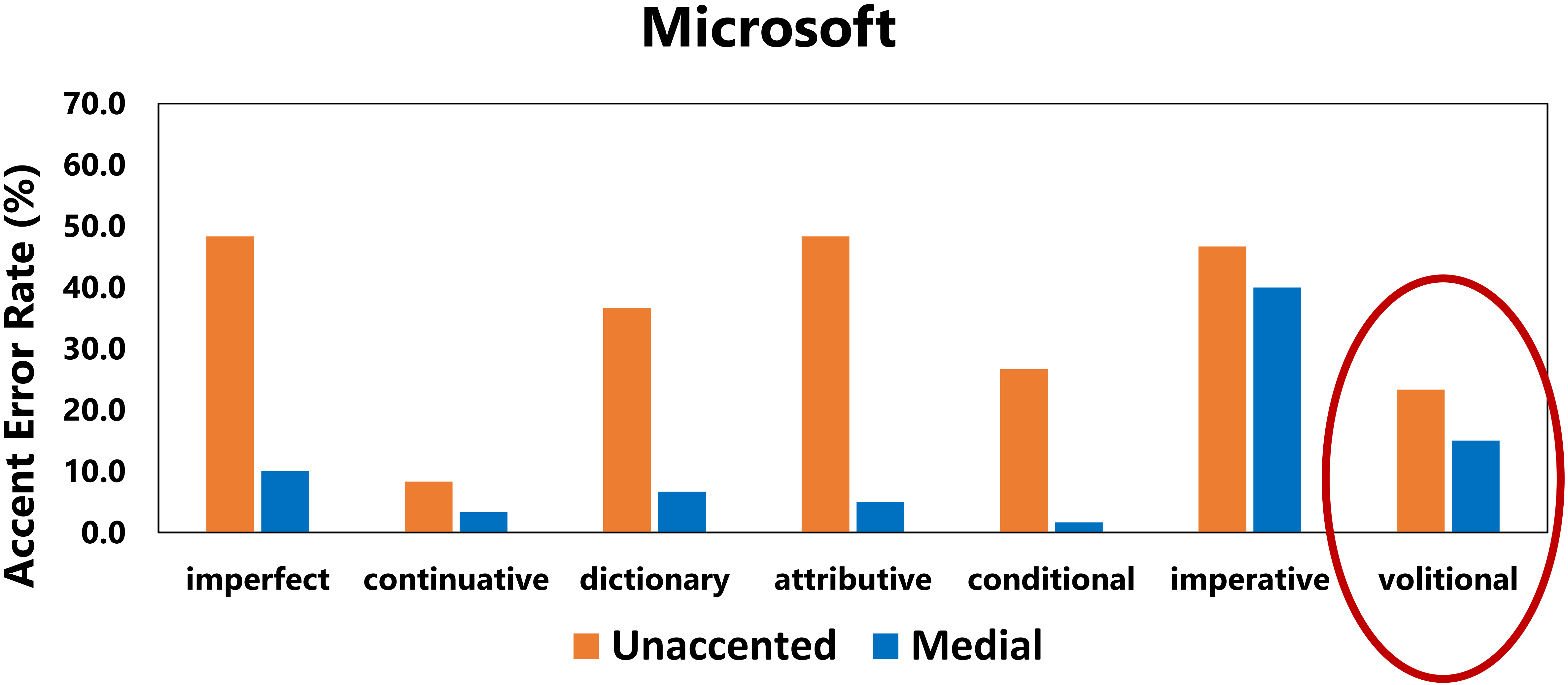


Figure7 : Accent error rate of Microsoft

Result

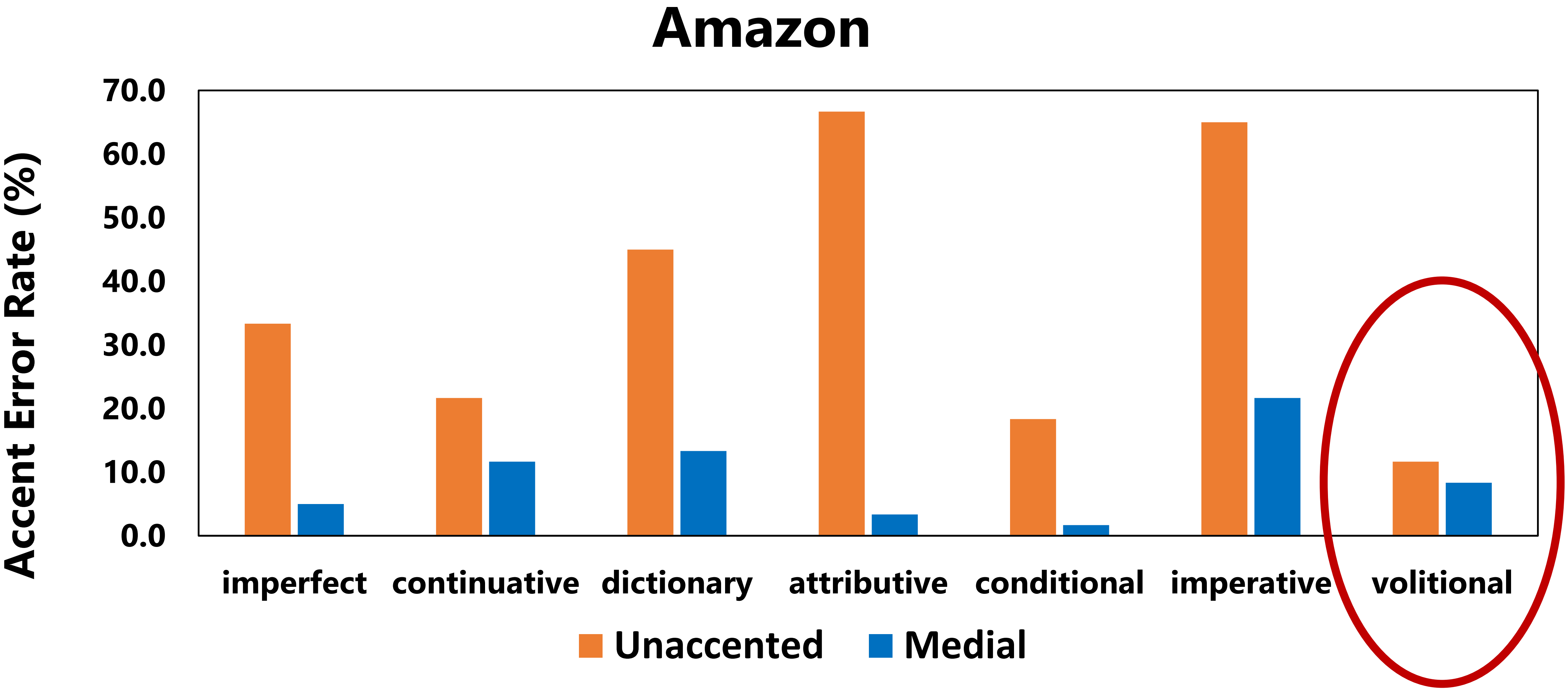


Figure8 : Accent error rate of Amazon

Common tendency 1

Microsoft

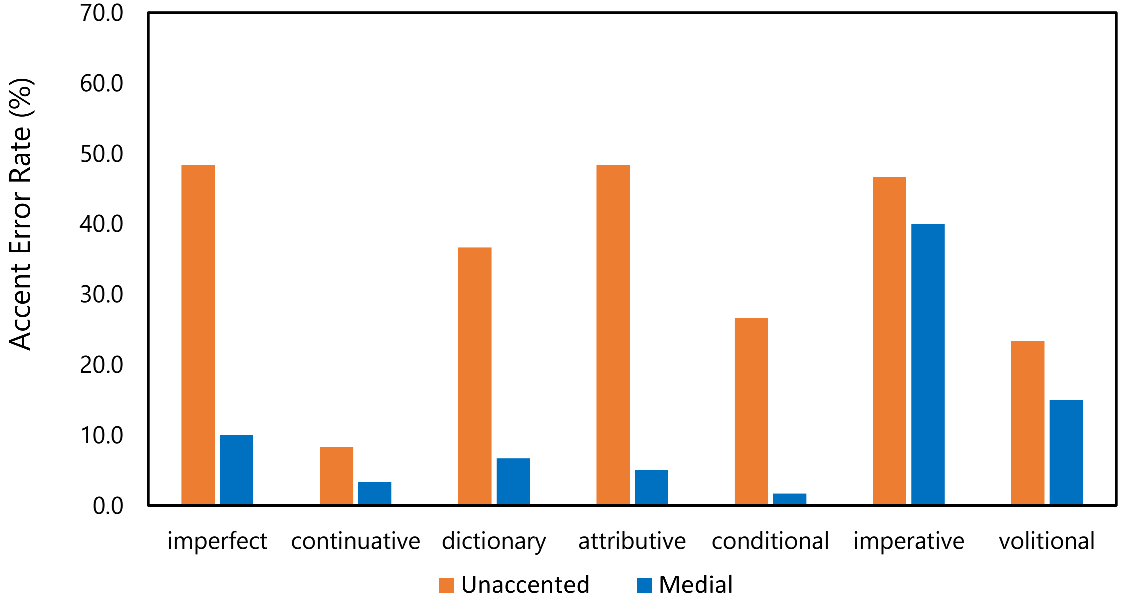


Figure7 (re): Accent error rate of Microsoft

Amazon

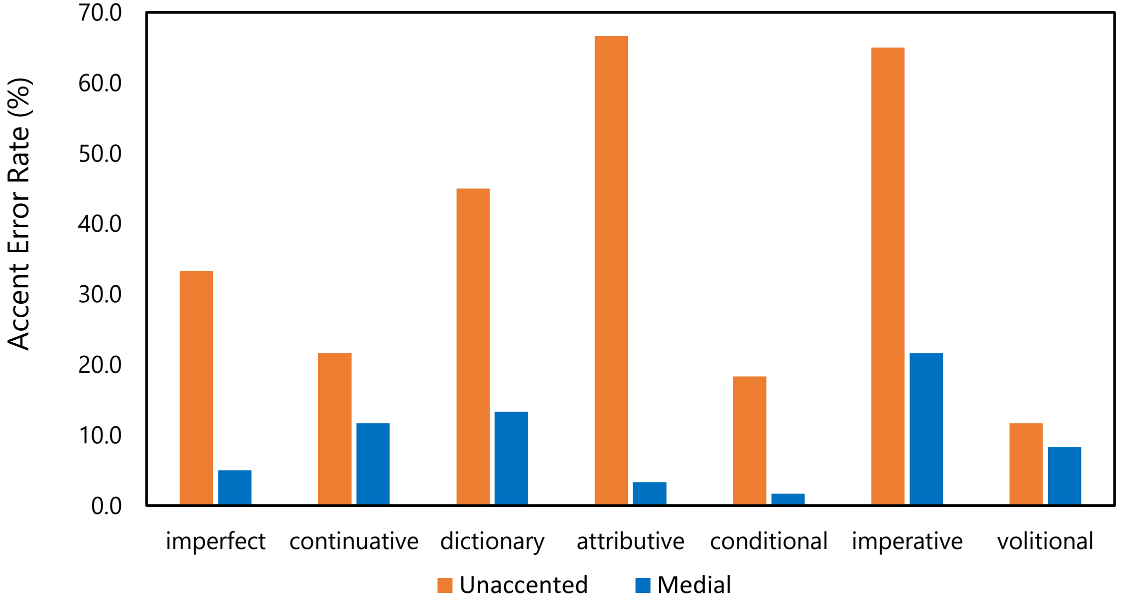


Figure8 (re): Accent error rate of Microsoft

- **Unaccented** had higher error rates than **Medial** accent

Common tendency 2

Microsoft

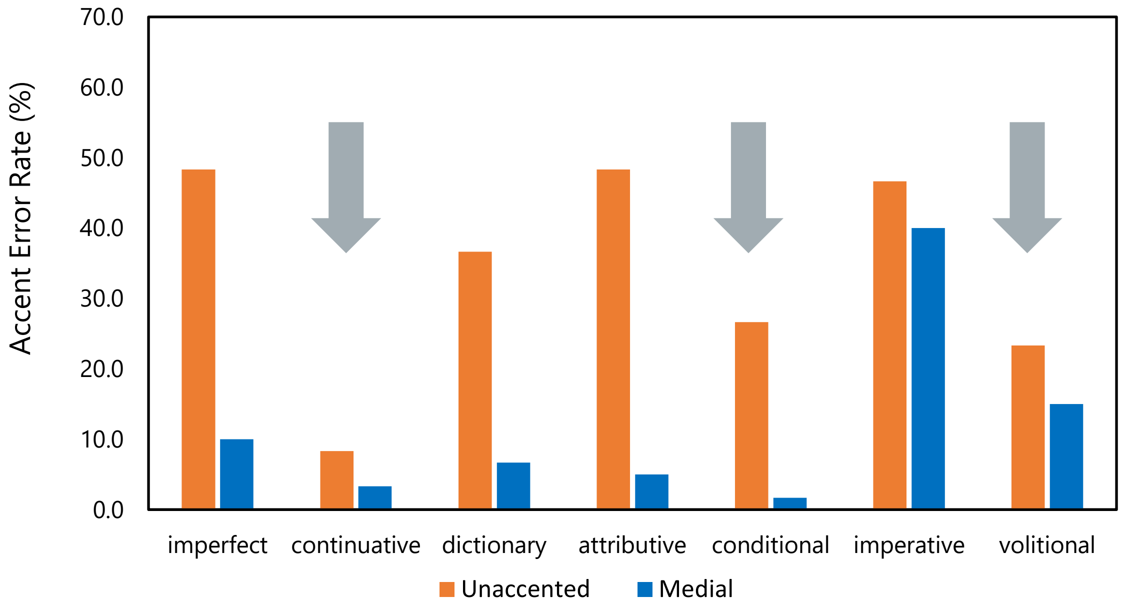


Figure7 (re): Accent error rate of Microsoft

Amazon

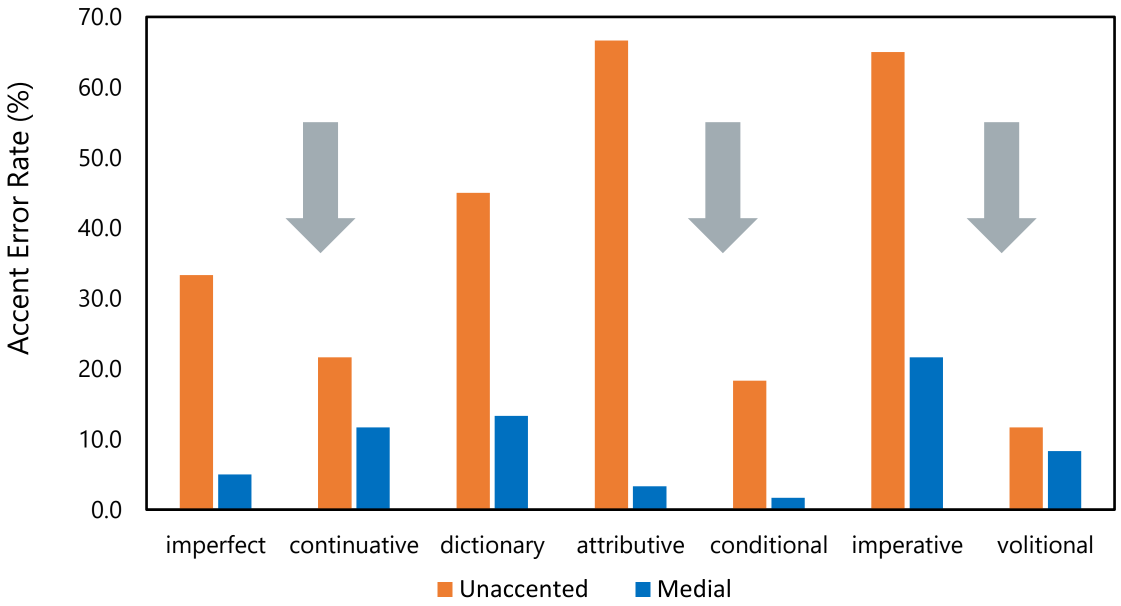


Figure8 (re): Accent error rate of Microsoft

- **Continuative, conditional, and volitional** were lower error rates **in unaccented**.

Discussion

Table 1 (re): The rules of accent variation [7]

	Imperfect	Continuative	Dictionary	Attributive
Unaccented	○○○-nai.	○○○-ma`su.	○○○.	○○○-hito.
Medial	○○○-`nai.	○○○-ma`su.	○○`○.	○○`○-tito.

	Conditional	Imperative	Volitional
Unaccented	○○○-`ba.	○○○.	○○○-masyo`u.
Medial	○○○-`ba.	○○`○.	○○○-masyo`u.

Conjugate form whose error rate was low had the accent nucleus.

➡ **Is the accent nucleus related to the error rate?**

[7] NHK Broadcasting Culture Research Institute (2016). “NHK nihongo hatsuon akusent sinjiten” (in Japanese). NHK syuppan

Outline

1. Introduction/Research Purpose
2. Study Plan
3. Results and Discussion
4. **Additional Experiment**
5. Conclusion and Future Study Plan
6. References
7. Question and Answer

Additional experiment 1

Remove accent nucleus by changing attached word

○○○-ma`su → ○○○-ta (both is continuative form)

Ex. Tabema`su → Tabeta

Result

●Both TTS's error rate was increased
(McNemar's test; both p < .01)

➡ No nucleus causes error!

Table2 : Error rate of additional experiment 1

	Before	After
Microsoft	8.30%	45.0%
Amazon	21.7%	56.7%

Additional experiment 2

Add accent nucleus by changing attached word

○○○-hito → ○○○-hito’he (both is attributive form)

Ex. Taberuhito → Taberuhito`he

Result

- **Both TTS’s error rate was decreased**
  (McNemar’s test; both p < .01)

➡ **Nucleus reduces error!**

Table3 : Error rate of additional experiment 2

	Before	After
Microsoft	48.3%	16.7%
Amazon	66.7%	20.%

Outline

1. Introduction/ Research Purpose
2. Study Plan
3. Results and Discussion
4. Additional Experiment
5. **Conclusion and Future Study Plan**
6. References
7. Question and Answer

Conclusion and Future Study Plan

High error rate

- Many unaccented
- When nucleus removed

→**They don’t have accent**

Low error rate

- Many medial accented
- Continuative, conditional, and volitional in unaccent
- When nucleus added

→**They have accent**

➡ **TTS tend to cause accent error in non-accent words.**

- We want to find the reason why no accent causes mistake.

References

- [1] iPhone Media. “Siri no uminooya ga kataru [genzai no Siri ni kaketeiru mono]” (in Japanese). https://iphone-mania.jp/news-205564/
- [2] Robosuta. “Sayonara Google home・Googme mini?” (in Japanese). https://robotstart.info/2020/05/28/google-home-no-longer-available.html.
- [3] piaro.net. “Character”. https://piapro.net/pages/character
- [4] Masayuki Suzuki, et al (2013). "Jokentukikakurituba wo motita Nihongo-Tokyohogen no akusentoketugo-jidosuitei" (in Japanese). The IEICE Transactions. D, 96(3), 644-654
- [5] Tadashi Sakamoto, et al (2017). "Nihongo-kyoiku heno michishirube dai 2 kan kotoba no shikumi wo shiru" (in Japanese). Bonjinsya
- [6] Kyoko Takeuchi, et al (2019). "Tanosi Onseigaku" (in Japanese). Kuroshiosyuppan
- [7] NHK Broadcasting Culture Research Institute (2016). “NHK nihongo hatsuon akusent sinjiten” (in Japanese). NHK syuppan
- [8] Microsoft. "Text to Speech". https://azure.microsoft.com/ja-jp/services/cognitive-services/text-to-speech/
- [9] Amazon. “Amazon Polly”. https://aws.amazon.com/jp/polly/
- [10] Nobuaki Minematsu, et al (2013). "Online Japanese Accent Dictionary". https://www.gavo.t.u-tokyo.ac.jp/ojad/

Thank you for your attention