"Hey Gūguru": Exploring Non-English Linguistic Barriers for Wake Word Use

Anna Bleakley[a], Yunhan Wu[a], Ayushi Pandey[b] and Justin Edwards[a]

[a]*University College Dublin (UCD), Belfield, Dublin 4, Ireland*
[b]*Trinity College Dublin, College Green, Dublin 2, Ireland*

Abstract
Wake words, or words that indicate the beginning of a command are fundamental to intelligent personal assistant (IPA) interactions across all users of IPAs. These commands, however, may vary greatly depending on the language of the user. As IPAs support more languages, it is important to consider how linguistic features vary and how these variations impact wake word usage. In this paper, we explore how linguistic features such as phonology, syntax, and pragmatics differ between languages and the implications of these differences on wake words in human-machine dialog design.

Keywords
Wake words, voice user interface, non-native language speakers, intelligent personal assistants,

1. Introduction

Wake words were initially introduced to intelligent personal assistant (IPA) interactions to overcome technical barriers and as an alternative to push-to-talk speech recognizers [1]. Today, they are used as a method of control for voice interactions [2]. When interacting with IPAs users use a single phrase to alert a device of a conversational turn, to indicate the device to listen. Previous research on wake words have shown that they are negatively received by users due to their relation to system errors [2], and other modalities are currently being explored to manage talk in conversation, as they impinge on the natural flow of conversation [3, 4].

As speech becomes a more common interaction modality, with an expanding number of supported languages [5], wake words need to be evaluated with considerations to linguistic features particularly in languages other than English. Recent findings [6, 7] have outlined that waking, turn-taking, and pronunciation are issues experienced by non-native language speakers when interacting with IPAs in English. These users therefore may have additional customized needs in supporting recognition of their speech. This becomes increasingly important as individuals that belong to households using multiple languages may be inclined to use IPAs in their non-native language. In addition to the increase of non-English speakers use of these interfaces as consumer markets for IPAs expand. Despite this, there is little work investigating how different languages impact IPA user experiences.

Currently, rigid phrases such as "Ok Google" or "Hey Alexa" are used as wake words, however they are not always transferable to other languages or to non-native speakers (L2) due to limitations in pronunciation, lexical knowledge, and cultural differences. For instance,

names used for IPAs are problematic for certain speakers to pronounce. Additionally greetings may be considered unusual expressions as precursors to commands. In this provocation, phonology, syntax, and pragmatics are used to explore considerations for future designs of wake words for IPAs with consideration of linguistic and cultural differences across a diverse and more global range of users.

2. Phonology

Language differences create ambiguities in identifying phonemes particularly in foreign words which may not have an equivalent in the speaker's language. Recognition issues therefore result from phoneme limitations for foreign words in a speakers' language. This problem is further exemplified in languages without clear lexical units, resulting in a problematic introduction in the use of English wake words. The Japanese alphabet contains 50 characters which are all phonemes in the Japanese language system. Due to phoneme limitations, Japanese speakers have a hard time pronouncing certain foreign words. For example, consonants [v] and [r] do not exist in the phonemic inventory of Japanese, speakers therefore create strategies to produce these words, producing the closest phonemic sounds [8]. This results in recognition issues with IPAs when using English expressions such as wake words.

Wake words result in similarly problematic issues in other languages such as Thai that do not have clear lexical units due to the absence of word boundary markers. In Thai, words are constructed with a combination of phonemes sometimes with low acoustic variability resulting in issues in automatic speech recognition (ASR) [9]. Furthermore, segmentation of morphemes is often dependent on individual speaker differences which creates difficulties for recognition due to unclear separation of lexical units. Foreign words particularly create awkward units and varied pronunciation among speakers as identified by [10]. For example, foreign names and transliterated English nouns do not have clear segments, resulting in unrecognizable pronunciation for speech systems. The following example of the name Mr. Robert from Jongtaveesataporn et al., [10] provides a case of the segmentation issue: นายโรเบิร์ต (Mr.Robert) is divided into นา (field) + ยโรเบิร์ต (unknown). The introduction of foreign names in wake words therefore creates problematic separations among users due to varied individual pronunciation differences as well as phonemic differences. To account for issues in relation to phonemes, wake words can integrate language specific information about phonemes to adopt guidelines on the recognition of wake words catered to available phonemes in different languages.

3. Syntax

Expressions such as "Hey Google" therefore provide familiar lexical access to speakers that are aligned with how utterances are planned in English reducing barge-in from IPAs [7]. In contrast, other languages with varying morphologies require different planning strategies due to varying linguistic typologies, particularly among L2 speakers. Especially in languages where the verb is at the end, for instance, Japanese has subject–object–verb (SOV) and sometimes object–subject–verb (OSV) word order. As a result, the projectability of a turn in Japanese might then take place at a later stage. Systems would then need to adjust how long they continue listening after a wake phrase to remedy different pause lengths after wake word phrases, particularly when lexical items introduced in wake word expressions are unusual in the speaker's native language. For example, phrases such as "Hey Alexa" are forced

expressions that may affect how a speaker continues their request when interacting with an IPA. As a result, users would likely struggle to project lexical items in their mother tongue after using a translated expression. Among L2 speakers, the planning of an utterance in their non-native language may require the systems to adjust the wait time for utterances accordingly due to resource limitations in processing [11]. L2 speakers therefore need more time to produce their utterances due to having to activate the language generation process in their non-native language. In human-machine dialog, L2 users generate whole sentences when interacting with IPAs to avoid recognition issues [7]. Hence, care must be taken when designing IPAs to accommodate utterances following wake words.

4. Pragmatics

For some languages, the use of an English product name like Alexa may present a distracting code switch between linguistic styles. Japanese has three types of alphabet systems, Hiragana, Kanji, and Katakana. Katakana is a language system which focuses on foreign languages. Foreign words are absorbed into the katakana system by pronouncing transliterated versions of the words using katakana characters in Japanese writing system [12]. Japanese also compounds nouns from foreign language into new katakana words to refer to a new meaning. Because of this borrowing and compounding strategy, Katakana expands very fast and fits Japanese people's general use. Although the Latin alphabet is contained in Japanese writing system, this alphabet is rarely used in daily life [12]. In this case, Japanese people don't usually use English words due to the powerful katakana system. Taking care that wake phrases may use katakana words in Japanese, even including

names for branded products like IPAs, would help prevent switching between disfavoured loanwords and favoured spoken forms when it isn't necessary. This is one way that unnatural use of foreign words may be avoided and may help to illustrate a general principle that languages may treat loanwords differently, challenging the norm of branded wake words by English language brands.

The very act of direct address may likewise be a pragmatic difference between English and other languages. Wake phrases most typically take the form of a verbal salute like "hey" or "okay" followed by the name of the IPA, leading all commands to begin with a direct address from the listener. Cultures vary across a number of dimensions including directness and individualism [13]. As such, cultures vary in terms of how much communication relies on context and mutual understanding as compared to how explicit communication is [14]. Cultures that are higher context may therefore use less direct address than explicit cultures as the intended listener would be understood by all communicators. English speaking North American countries, where most IPAs are made, have low context cultures which use very explicit communication [15]. The use of the vocative in a wake phrase may thus be a less natural experience for IPA users from higher context cultures where this sort of address would not be used as its social function would ordinarily be captured by context. Understanding the ways different cultures use direct address can help IPA designers plan for cultural expectations and norms, allowing IPA interactions to achieve comparable degrees of naturalness across culture.

5. Conclusion

Interactions with IPAs fundamentally require the use of specific phrases or behaviours, wake

phrases, that initiate interactions. These phrases are currently monolithic in their form and word choice, including greetings and direct vocative address of the IPA. The lack of choice and diversity in wake phrases may however lead to exclusion or feelings of unnaturalness for users who speak different languages or are made to interact with IPAs in a non-native language. If IPAs are to include users who speak languages other than English, care must be taken to select and allow for wake phrases that accommodate different linguistic abilities and expectations. Considerations about how languages differ in terms of phonology, command syntax, and pragmatics of direct address all must be made to ensure that the experience of using an IPA is comparable for all users.

Acknowledgments

References

[1] V. Këpuska, Wake-up-word speech recognition, Speech Technologies (2011) 237–262.

[2] H. Jung, H. Kim, Finding contextual meaning of the wake word, in: Proceedings of the 1st International Conference on Conversational User Interfaces, 2019, pp. 1–3.

[3] D. McMillan, B. Brown, I. Kawaguchi, R. Jaber, J. Solsona Belenguer, H. Kuzuoka, Designing with gaze: Tama–a gaze activated smart-speaker, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–26.

[4] Z. Yang, C. Yu, F. Zheng, Y. Shi, Proxitalk: Activate speech input by bringing smartphone to the mouth, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3 (2019) 1–25.

[5] B. Kinsella, Google assistant now supports simplified chinese on android smartphones, https://voicebot.ai/2019/05/15/google-assistant-now-supports-simplified-chinese/, 2020.

[6] Y. Wu, J. Edwards, O. Cooney, A. Bleakley, P. R. Doyle, L. Clark, D. Rough, B. R. Cowan, Mental workload and language production in non-native speaker ipa interaction, in: Proceedings of the 2nd Conference on Conversational User Interfaces, 2020, pp. 1–8.

[7] Y. Wu, D. Rough, A. Bleakley, J. Edwards, O. Cooney, P. R. Doyle, L. Clark, B. R. Cowan, See what i'm saying? comparing intelligent personal assistant use for native and non-native language speakers, in: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, 2020, pp. 1–9.

[8] K. Aoyama, J. E. Flege, S. G. Guion, R. Akahane-Yamada, T. Yamada, Perceived phonetic dissimilarity and l2 speech learning: The case of japanese/r/and english/l/and/r, Journal of Phonetics 32 (2004) 233–250.

[9] G. Saon, M. Padmanabhan, Data-driven approach to designing compound words for continuous speech recognition, IEEE transactions on Speech and audio processing 9 (2001) 327–332.

[10] M. Jongtaveesataporn, I. Thienlikit,

C. Wutiwiwatchai, S. Furui, Lexical units for thai lvcsr, Speech Communication 51 (2009) 379–389.

[11] Z. Dörnyei, J. Kormos, Problem-solving mechanisms in l2 communication: A psycholinguistic perspective, Studies in second language acquisition (1998) 349–385.

[12] N. Kaji, M. Kitsuregawa, Splitting noun compounds via monolingual and bilingual paraphrasing: A study on japanese katakana words, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 959–969.

[13] G. Hofstede, Culture's consequences: International differences in work-related values, volume 5, sage, 1984.

[14] E. T. Hall, Beyond culture, Anchor, 1989.

[15] S. Nishimura, A. Nevgi, S. Tella, Communication style and cultural features in high/low context communication cultures: A case study of finland, japan and india, Teoksessa A. Kallioniemi (toim.), Uudistuva ja kehittyvä ainedidaktiikka. Ainedidaktinen symposiumi 8 (2008) 783–796.