

How Long is Too Long?: Excessive Pause Duration in Voice User Interfaces

Lucas J. Hess and Daniela Barron

Introduction

- What is a VUI?
 - A Voice User Interface (VUI) is what an individual interacts with in a spoken language application in order to accomplish a task or receive assistance, such as Apple's "Siri" or automated attendants in information technology support
 - The design of a VUI's grammar, promptness, prosody, and call flow is important to its design
- When interacting with VUI's, humans have perceptual temporal expectations of this interaction which result from experiences (Cohen et al., 2004)
 - Such as pauses and gaps in human-to-human conversation



Introduction

- When these expectations are violated, this results in interfaces that are perceived as:
 - Less comfortable
 - Having less flow
 - More difficult to interact with and comprehend
- Temporal perceptual expectations as to when a VUI should provide feedback and respond to their prompts may result in negative user experience effects (Commarford & Lewis, 2005).



Goal

- At what point is too long of a pause between when a human speaker provides input and a Voice User Interface (VUI) responds?



Research Utilized

- Revealed that the research on human temporal perception of a VUIs voice feedback was scarce
- Our research also incorporated research in human-human conversation
- By utilizing norms in human-human conversation, we can understand the perceptual expectations of the users related to VUI interactions (Gravano & Hirschberg, 2011)
- Focus on research which utilized telephone conversation and had no eye contact interactions to control for non-speech cues



Average Durations

- Range from Research:
 - 100-500 ms for inter-speaker pause durations
- Average Pause Duration:
 - ~ 300-350 ms
- Provides reference for how long typical pause durations tend to be, so that we can understand the difference between average pause durations compared to max durations
- We can estimate a normal conversational or VUI experience (Gravano & Hirschberg, 2011).
- Understand where a noticeable difference in temporal feedback is perceptually apparent and undesirable to the user



Table 1: Average Durations of Inter-Speaker Pauses Indicated by Past Research:

Authors	Duration (ms)	Relevant Summary
Baumann (2008)	331-363	Investigated turn-taking strategies in a simulated environment. Participants exchanged audio streams in real-time, and autonomously judge turn-taking behavior.
Beattie & Barnard (1979)	250	Investigated timing of turn taking during American English service-based conversations over the phone.
Brady (1968)	345-456	Investigated gaps in sixteen phone calls between friends in the USA.
Gravano (2009)	100-200	Investigated the final and initial utterances of turns in a conversation using task-oriented dialog, and ways to potentially predict what kind of turn-yielding cues someone might be using for applications into VUI systems.
Holler et al. (2016)	100-500	Analyzed human-to-human conversational structure, and presented an overview of the research and literature surrounding turn-taking in conversations.

Table 1: Average Durations of Inter-Speaker Pauses Indicated by Past Research Part 2:

Kendrick & Torreira (2015)	300	Indicated from corpus analysis that gaps longer a norm of 300 ms decrease likelihood of an unqualified acceptance and dispreferred turn format.
Norwine & Murphy (1938)	410	Investigated pauses in calls on a New York-Chicago telephone circuit used for Bell System business.
Sellen (1995)	480	Videoconferencing systems were evaluated experimentally and differed based on if participants were visible and if they were in the same room. The duration included is when speakers were not visible to each other.
Stivers (2009)	200	Investigated universal basis for turn-taking behavior demonstrated between all languages studied.
Weilhammer & Rabold (2003)	380	Investigated task-oriented telephone conversation and pauses between English, German, and Japanese. The mean reported is English speakers.
Wilson & Wilson (2005)	110-400	Investigated using brain oscillation as a technique to understand turn-taking in conversation and delves into how a speaker and a listener can become entrained by identifying rate of speech and syllable production.

Max Durations: What We Want to Know


- Average max pause time duration indicated by previous research:
 - ~ 1000 ms (1 second), ± 100 ms
- Max duration of inter-speaker pauses within human to human conversations and VUI interaction.
- Summarized available research that can indicate an undesirable duration
- Range:
 - 500-1300 ms



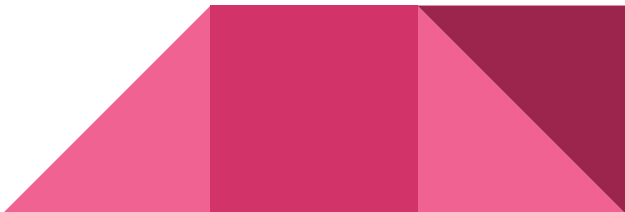
Table 2: Max Durations of Inter-Speaker Pauses Indicated by Past Research:

Author	Duration (ms)	Relevant Summary
Beattie & Barnard (1979)	1250	Investigated temporal characteristics of speaker transitions in natural telephone conversation.
Commarford & Lewis (2005)	1300	Presented analysis on optimal pause duration between menu presentation and global navigation commands in a VUI system.
Heldner and Edlund (2010)	500-1000	Explored durational aspects of pauses, gaps, and overlaps in conversational corpora for use in speech technology design.
Kendrick & Torreira (2015)	700	Corpus Analysis demonstrated that gaps longer than 700 ms indicated negative effects.
Roberts et al. (2011)	600	Universal temporal mechanisms of spoken language were investigated using telephone conversations between friends.
Wilson and Wilson (2005)	910-1,000	Investigates using brain oscillation as a technique to understand turn-taking in conversation, and delves into how a speaker and a listener can become "mutually entrained through recognizing rate of speech and syllable production."

Negative Effects of Excessive Pause Durations

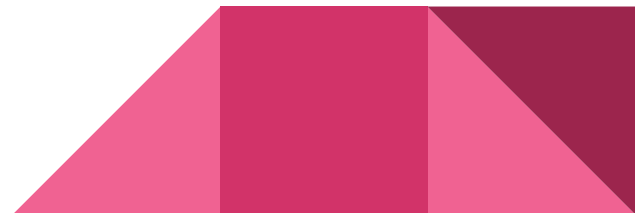
- Perceived Excessive pause durations can generate negative effects to the users experience (Beattie & Barnard, 1979; Commarford & Lewis, 2005)
 - Such as:
 - Negative inferences about increase in likelihood of a dispreferred response (Roberts et al., 2011)
 - Perceived decreased likelihood of a preferred response (Kendrick and Torreira, 2015)
 - User may want to speak again after a certain time (Beattie & Barnard, 1979; Commarford & Lewis, 2005).
 - Perception of decreased likelihood of a general acceptance and increased the likelihood that a response will possess a dispreferred turn format, such as “Yes...but” (Kendrick and Torreira, 2015)
 - Make users uncomfortable (Cohen et al., 2004)
 - Can result in users making inferences about the upcoming responses and respond out of turn (Beattie & Barnard, 1979; Commarford & Lewis, 2005; Roberts et al., 2011)
- 

Limitations: Factors that Contribute to Variability in Perception of Pause Duration:

- Large variability in research for average and max pause durations may be because:
 - Individuals tend to match their new conversational partners in terms of pause durations (ten Bosch et al., 2004, 2005; Wilson and Wilson 2005; Gravano 2009).
 - Different turn-yielding cue combination typically warrants a different turn-taking interval (Gravano, 2009)
 - Longer a speaker plans to speak, more cognitive preparation is needed to produce longer responses and warrants longer pauses (Torreira et al., 2015)
 - Differences in methodologies in research utilized
- 

Limitations: Factors that Contribute to Variability in Perception of Pause Duration:

- Inter-conversational pause durations vary for different languages.
 - English (380 ms), German (363 ms), and Japanese (389 ms) (Weilhammer & Rabbold, 2003).
- Other research has argued that this is negligible difference and noted that the factors that affect response times are often similar, regardless of language or culture (De Ruiter et al. 2006; Levinson et al., 2015; Norwine and Murphy 1938; Sellen 1995; Stivers et. al. 2009)



Conclusions

- Max (Excessive) Pause Duration
 - ~1000 ms (1 second), ± 100 ms
- Average (Normal) Pause Duration:
 - ~300-350 ms
- We provide estimate of excessive pause durations and the negative effects associated with late conversational and VUI feedback
- Designers should attempt to not exceed Max Pause Duration or negative effects to user experience may ensue




Implications

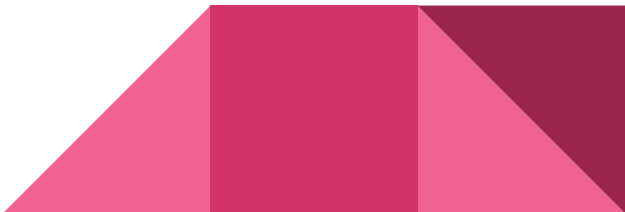
- Future research should investigate a more exact duration threshold for excessive pauses in human-machine voice interaction and VUI design
- Negative effects mentioned can affect business
 - Such as increase in errors due to conversational turn overlap
- The importance of implementation of the duration cap of 1000 ms, or 1 second (± 100 ms) into operational settings is vital for the user's experience, general function of the VUI, and consequently, the success of the business




References

- Baumann, T. (2008). Simulating spoken dialogue with a focus on realistic turn-taking. Proceedings of the 5th International Workshop on Constraints and Language Processing, 1-8.
- Beattie, G. W., & Barnard, P. J. (1979). The temporal structure of natural telephone conversations (directory enquiry calls). *Linguistics*, 17(3-4), 213-230.
- Brady, P. T. (1968). A statistical analysis of on-off patterns in 16 conversations. *Bell System Technical Journal*, 47(1), 73-91.
- Cohen, M. H., Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. Addison-Wesley Professional.
- Commarford, P. M., & Lewis, J. R. (2005). Optimizing the pause length before presentation of global navigation commands. In *Proceedings of HCI*, 2, 1-7.
- 

References Cont.

- Gravano, A. (2009). Turn-taking and affirmative cue words in task-oriented dialogue. Columbia University Computer Science Technical Reports. *Department of Computer Science, Columbia University*, CUCS-009-09, 1-219.
- Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3), 601-634.
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555-568.
- Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2016). *Turn-taking in human communicative interaction*. Frontiers Media.
- Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4), 255-289.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- 

References Cont.

- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, 6, 731.
- Norwine, A. C., & Murphy, O. J. (1938). Characteristic time intervals in telephonic conversation. *Bell System Technical Journal*, 17(2), 281-291.
- Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, 48(5), 331-354.
- Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-computer interaction*, 10(4), 401-444.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587-10592.
- 



Questions?