VISION

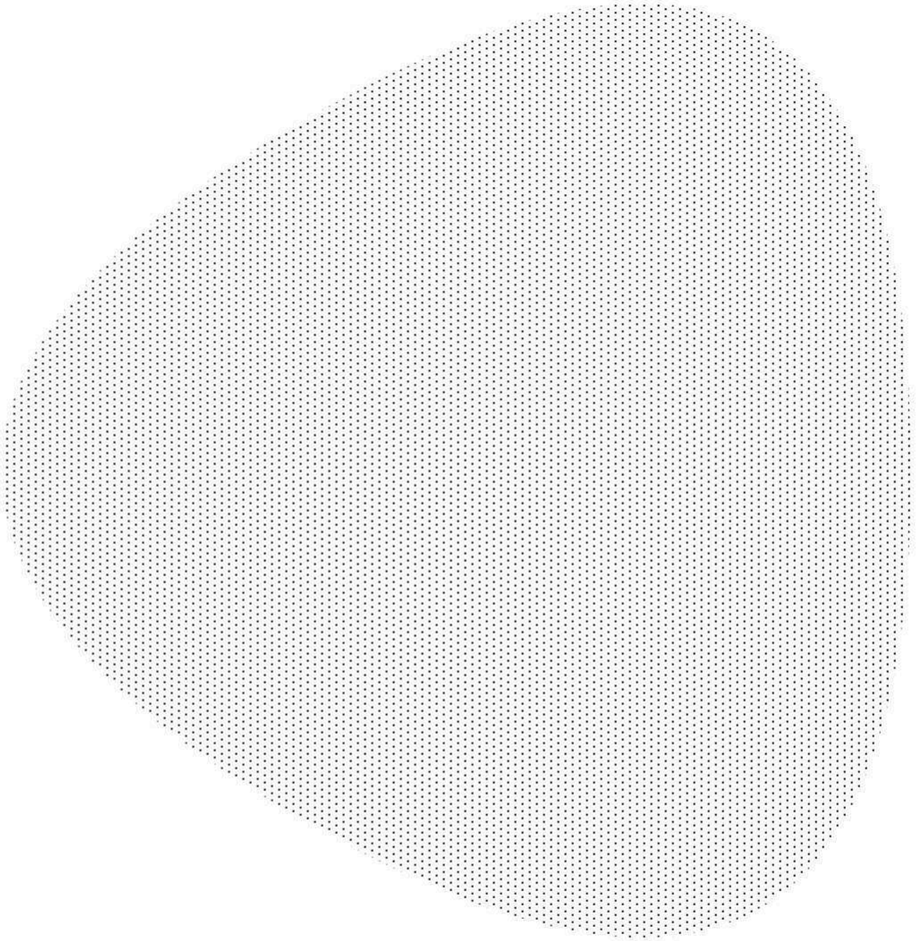
Making talking to an Agent
more natural

GROUNDING IN COMMUNICATION WITH AN A.I.

8 week Internship Project
May-July 2019
Anshuman Dhar / IIT Guwahati
Voice UX Team

Grounding in Communication

The untaught rules of communication which all humans have.



What is more natural?

I looked into the different areas which could make an agent more natural to talk to.

MIRRORING IN CONVERSATION

CONVEY EMOTIONS

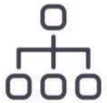
HOW THE AGENT SOUNDS/ UNDERSTANDS YOU

DYNAMIC NOISE LEVELS

PAUSES IN CONVERSATION

TALK TO MULTIPLE USERS

INTERACTION ACROSS DEVICES



INVOCATION

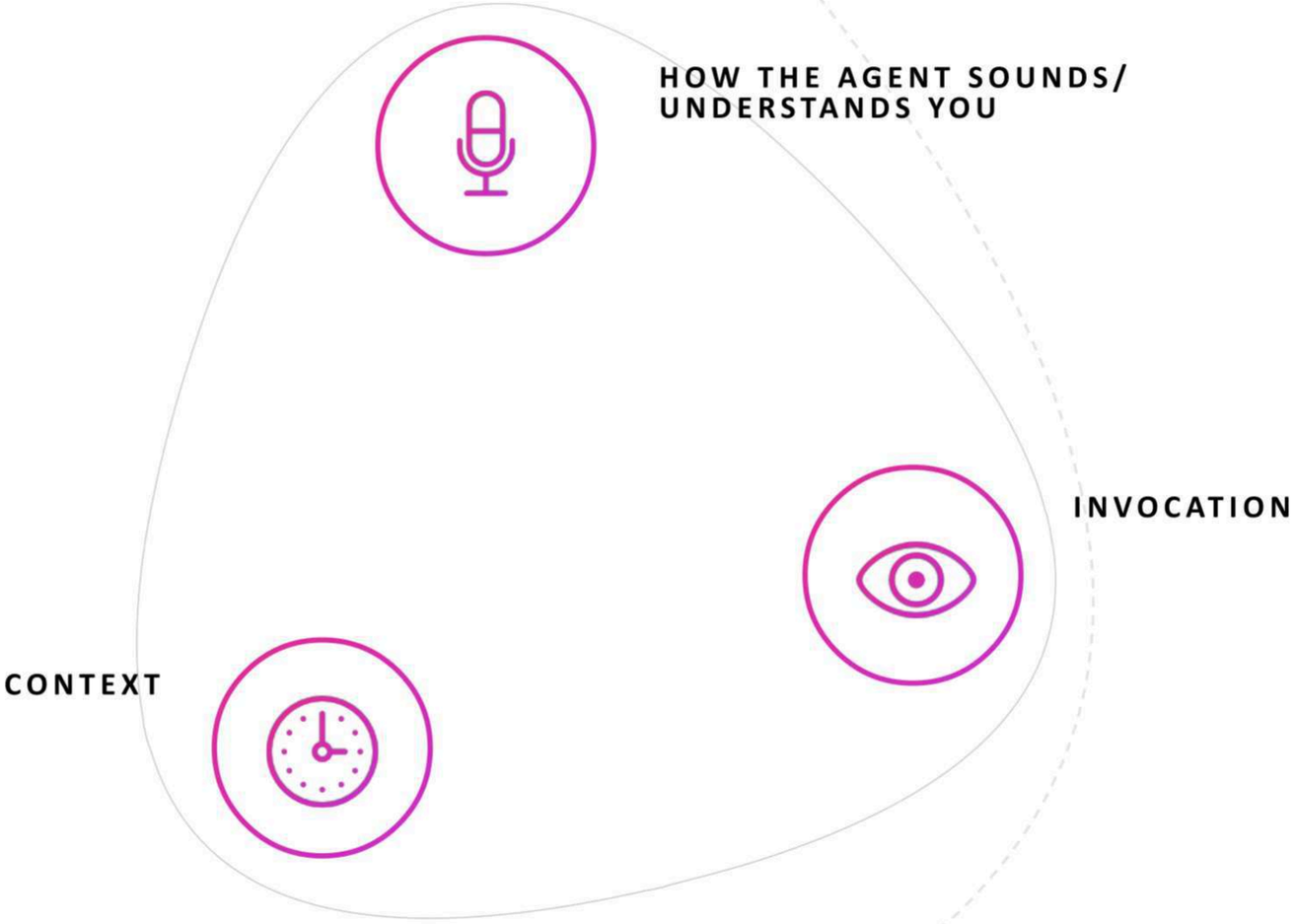
LOWER LATENCY

CONTEXT

GESTURE BASED INTERACTIONS

PROACTIVE INTERRUPTIONS

What is more natural?



What is more natural?

HOW THE AGENT SOUNDS/ UNDERSTANDS YOU

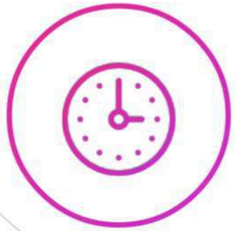
Looking at ASR and TTS and understanding how they work.

INVOCATION

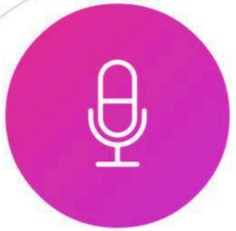
Looking at better ways to start and sustain conversation.

CONTEXT

Figuring out how the agent can be more aware of the context user is in.



What is more natural?



HOW THE AGENT SOUNDS/ UNDERSTANDS YOU

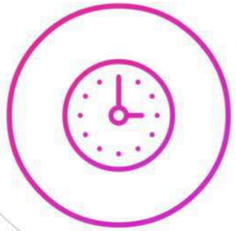
Looking at ASR and TTS and understanding how they work.

INVOCATION

Looking at wake words and the different kinds of ways to initiate dialogue and sustain conversation.

CONTEXT

Figuring out how the agent can be more aware of the context user is in.



NATURAL LANGUAGE UNDERSTANDING (NLU)

How the Agent Understands


Statistical natural language understanding using hidden clumpings

Abstract

Images (2)


Transportable Natural Language Processing through Simplicity—The PRE System

SAMUEL S. EPSTEIN
Gel Communications Research

1. INTRODUCTION

Extensive Literature Review

How does the Agent know if ‘foot’ is a measurement or a body part?


Spoken Language Understanding for Natural Interaction: The Siri Experience

Abstract

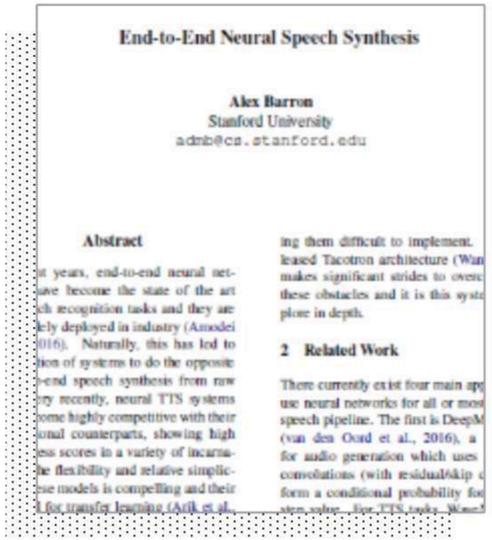
Keywords

TEXT TO SPEECH (TTS)

What the Agent Sounds like

How does an agent know how

to vocalize words/letters?



End-to-End Neural Speech Synthesis

Alex Barron
Stanford University
admb@cs.stanford.edu

Abstract

2 Related Work

Characterization and Incorporation of Emotions in Speech

T. V. Sagar, K. S. Rao, S. R. M. Prasanna, and S. Dandapat

Sounding more natural

EMOTION

Expressed through the timbre of voice and the mean pitch value

PROSODY

Collection of things, such as intonation, stress, and rhythm.

PAUSES

Hmms, aahs and fake pauses are added when processing takes too long.

Sounding more natural

All the things I read about human communication etc how an agent can be made to sound even more like a human being.

PIVOT

INTONATION

There are many variations of passages of Lorem Ipsum available, but the majority.

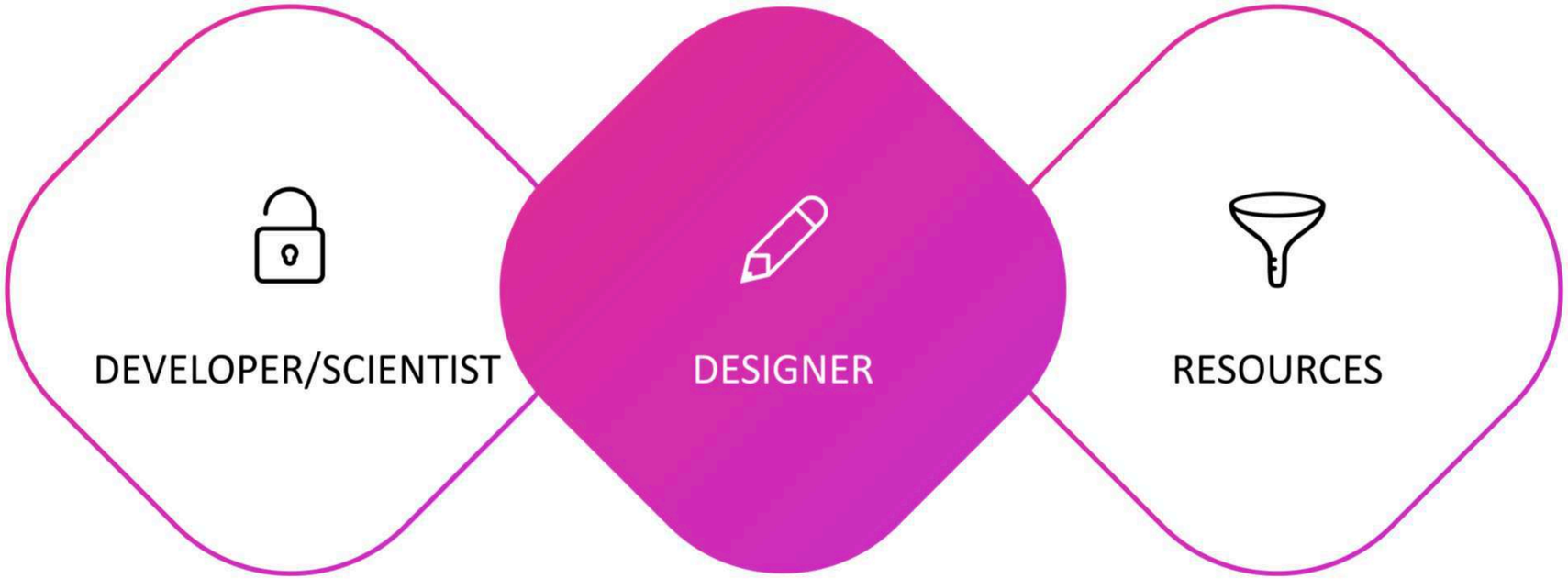
PROSODY

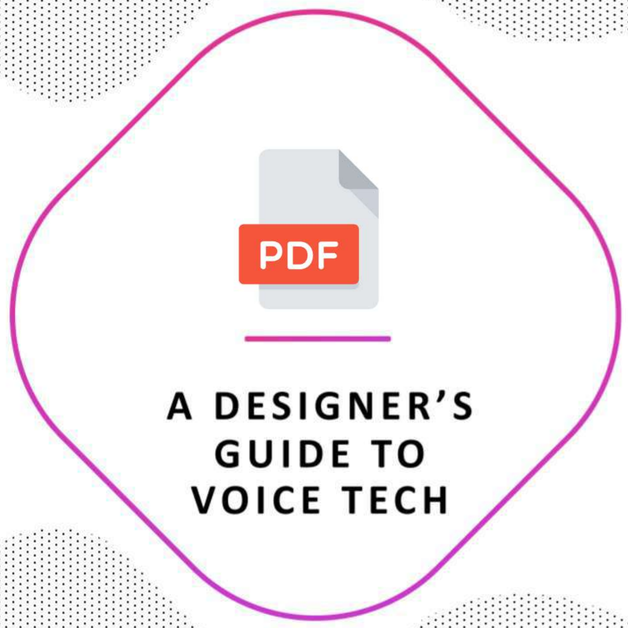
There are many variations of passages of Lorem Ipsum available, but the majority.

PAUSES

There are many variations of passages of Lorem Ipsum available, but the majority.

Why Pivot





Blog/Research

Did anything productive come out of this?

A simple 101 guide for how agents speak and understand a language

Features

Understanding

- Automatic Speech Recognition
- Indexing

Speaking

- Machine Learning
- Neural Networks
- Recurrent Neural Networks
- Long Short Term Memory
- Text-To-Speech (TTS)
- Concatenative TTS
- End-to-End Neural TTS

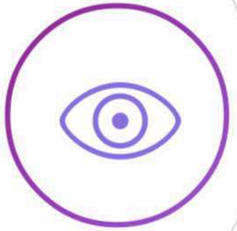
Microsoft
werPoint Presentat

What is more natural?

UPDATE: A Blog/ Guide for designers to understand voice tech

HOW THE AGENT SOUNDS/ UNDERSTANDS YOU

Looking at ASR and TTS and understanding how they work.



INVOCATION

Looking at better ways to start and sustain conversation.

CONTEXT

Figuring out how the agent can be more aware of the context user is in.

PROACTIVE INTERRUPTION
(CONTEXTUAL AWARENESS)

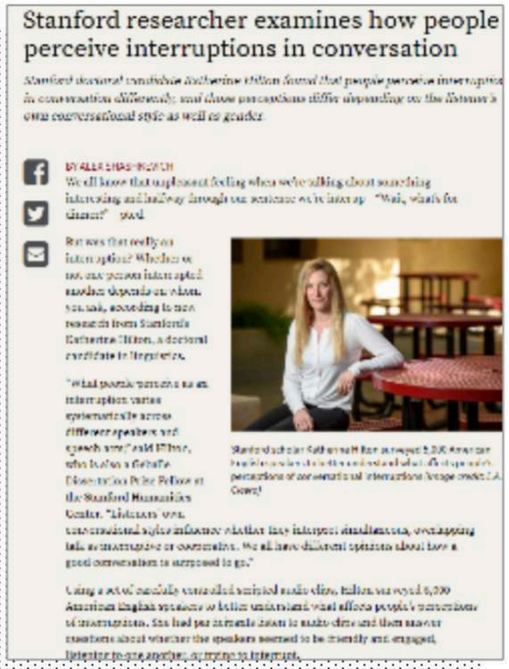
When should the Agent interrupt?

Hang on a Sec! Effects of Proactive Mediation of Phone Conversations while Driving

Shamsi T. Iqbal[1], Eric Horvitz[1], Yun-Cheng Ju[1] and Ella Mathews[2]

[1]Microsoft Research
Microsoft Way, Redmond, WA 98052

[2]California Institute of Technology
1200 E California Boulevard, Pasadena, C



Stanford researcher examines how people perceive interruptions in conversation

Stanford doctoral candidate Katherine Ellison found that people perceive interruptions in conversation differently, and those perceptions differ depending on the listener's own conversational style as well as gender.

BY ALLEE EINSPRUCH

We all know that unpleasant feeling when we're talking about something interesting and halfway through our sentence we're interrupted. "Wait, what's for dinner?" you'd

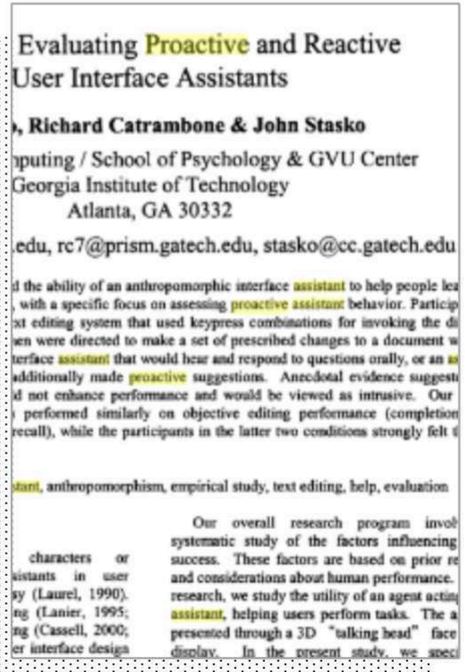
But was that really an interruption? Whether or not the person interrupted another depends on whom you ask, according to new research from Stanford's Katherine Ellison, a doctoral candidate in linguistics.

"What people perceive as an interruption varies systematically across different speakers and speech acts," said Ellison, who is also a Graduate Fellow at the Stanford Humanities Center. "Listeners' own conversational styles influence whether they interpret simultaneous, overlapping talk as interruptive or cooperative. We all have different opinions about how a good conversation is supposed to go."

Using a set of carefully controlled recorded audio clips, Ellison surveyed 6,700 American English speakers to better understand what affects people's perceptions of interruptions. She had participants listen to audio clips and then answer questions about whether the speakers seemed to be friendly and engaged, listening to one another, or trying to interrupt.

Extensive Literature Review

How can an agent know when the best time to interrupt is?



Evaluating Proactive and Reactive User Interface Assistants

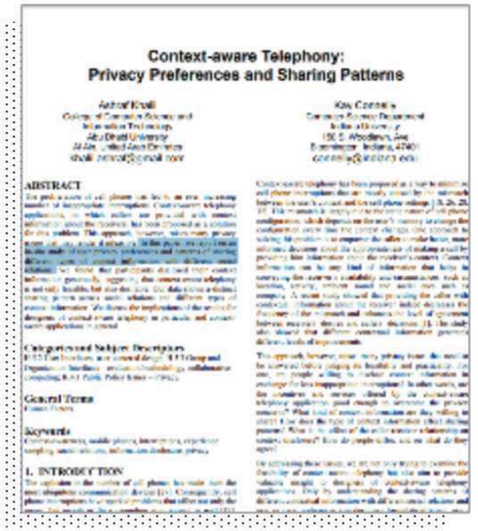
, Richard Catrambone & John Stasko

nputing / School of Psychology & GVU Center
Georgia Institute of Technology
Atlanta, GA 30332

.edu, rc7@prism.gatech.edu, stasko@cc.gatech.edu

PRIVACY CONCERNS (DATA RECORDED)

What are the users fine with sharing?


Context-aware Telephony:
Privacy Preferences and Sharing Patterns

ABSTRACT

Would a user be okay with giving their location data if it meant getting more relevant notifications?

ABSTRACT

The proliferation of cell phones has led to an ever increasing number of inappropriate interruptions. Context-aware telephony applications, in which callers are provided with context information about the receivers, has been proposed as a solution for this problem. This approach, however, raises many privacy issues that may render it infeasible. In this paper, we report on an in-situ study of user privacy preferences and patterns of sharing different types of context information with different social relations. We found that participants disclosed their context information generously, suggesting that context-aware telephony is not only feasible, but also desirable. Our data shows a distinct sharing pattern across social relations and different types of context information. We discuss the implications of the results for designers of context-aware telephony in particular and context-aware applications in general.

Categories and Subject Descriptors

H.5.2 User Interfaces- user-centered design, H.5.3 Group and Organization Interfaces – evaluation/methodology, collaborative computing; K.4.1 Public Policy Issues – Privacy.

General Terms

Human Factors.

Keywords

Context-awareness, mobile phones, interruptions, experience sampling, social relations, information disclosure, privacy.

Being *aware* of context

PROACTIVE ASSISTANCE

Proactively offering help to the user at the right point can be a huge help

PRIVACY PREFERENCES

Depends of the context of use.

INTERRUPTIONS

Badly timed interruptions can be extremely harmful.

Concept

A new kind of Agent

An assistant in a Multi-Device Environment (MDE) which understands the context you are in.

Context-Aware Agent

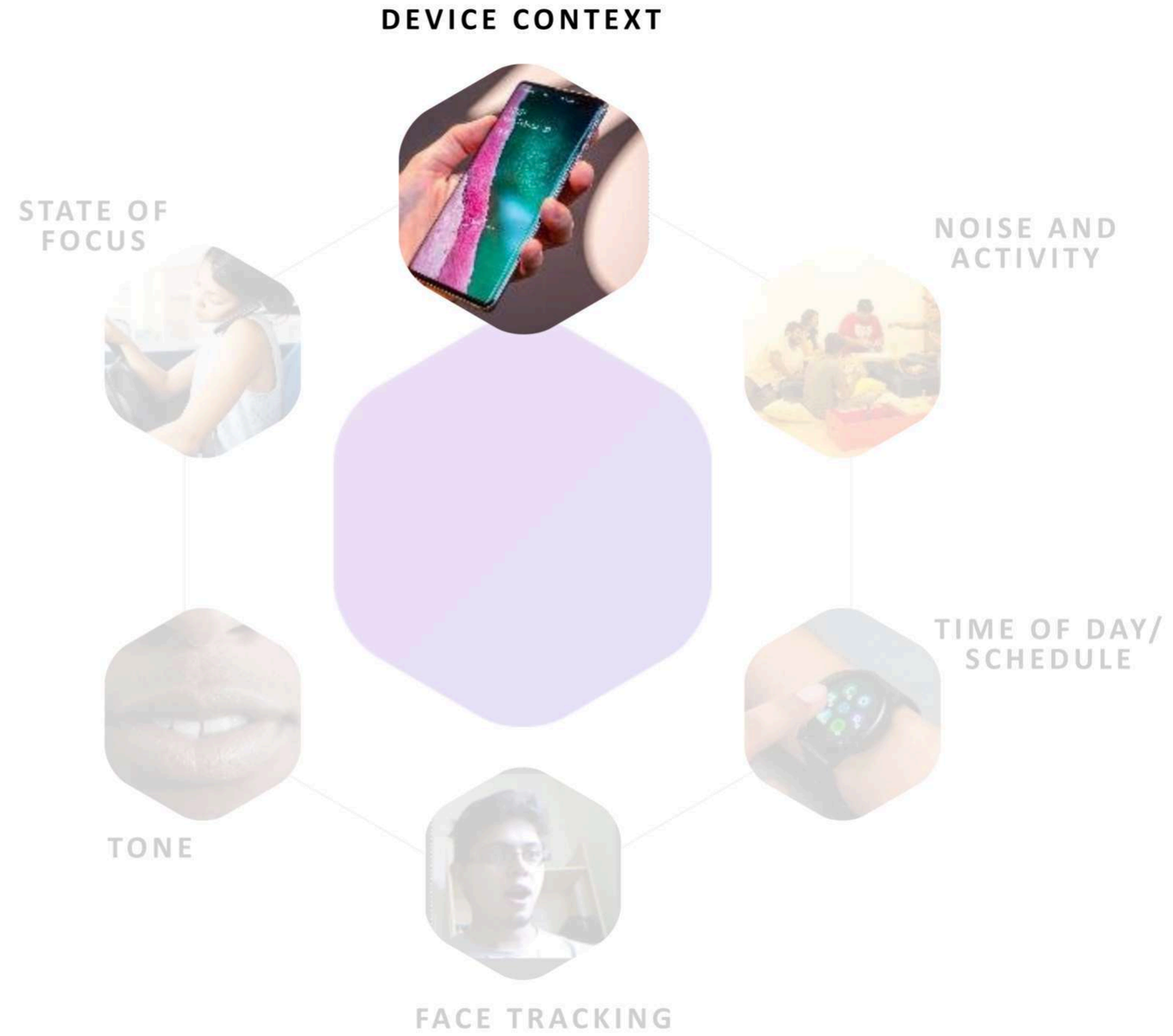
In an MDE., a combination of these parameters will be taken by the agent.

*Multi Device Environment

User was using the phone, then kept it on the table.

User asks agent to send them their flight tickets.

Agent recognizes that the last device used by user was phone and sends it there

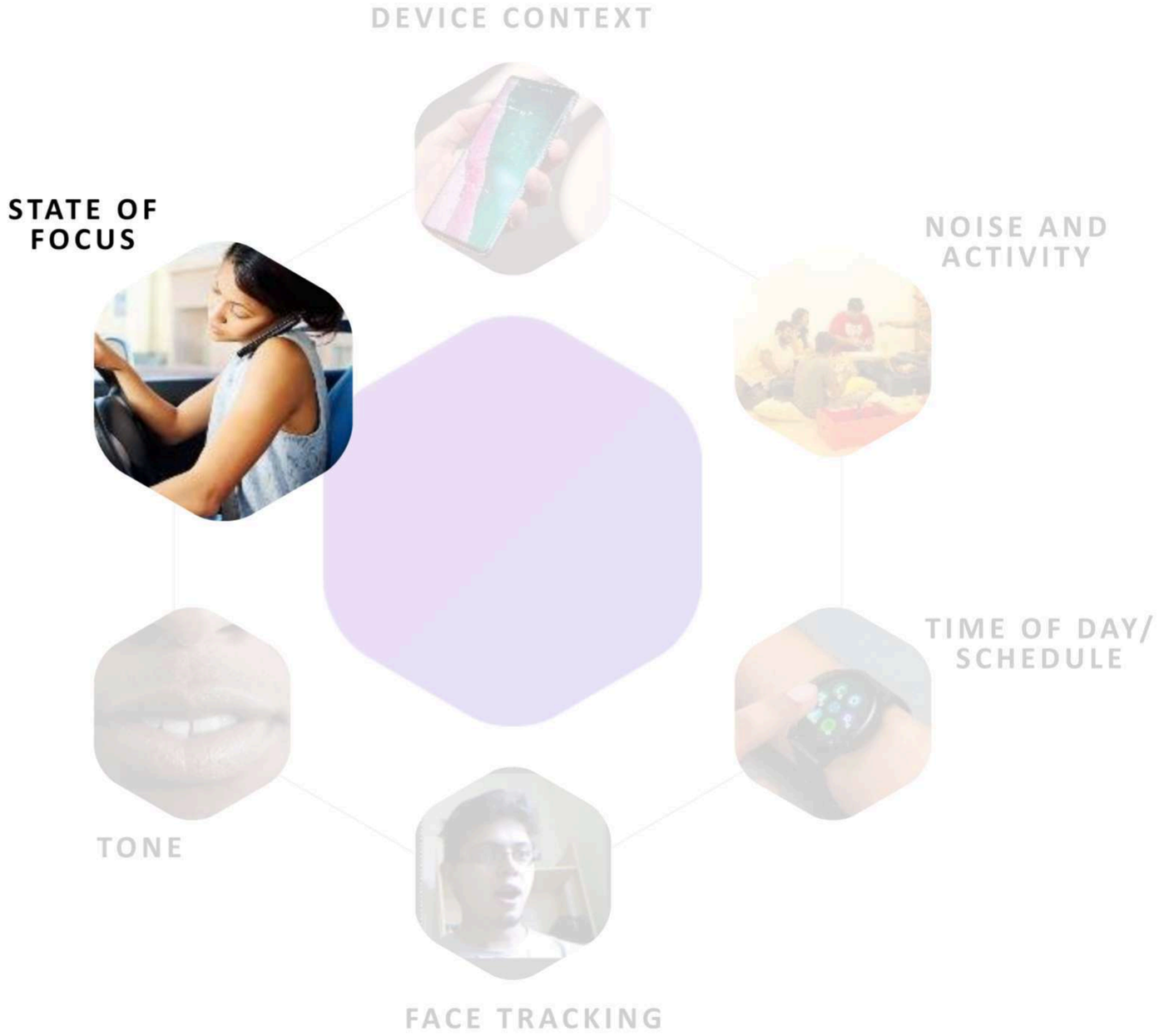


*Multi Device Environment

User is making a long presentation/report

User has previously asked agent to notify

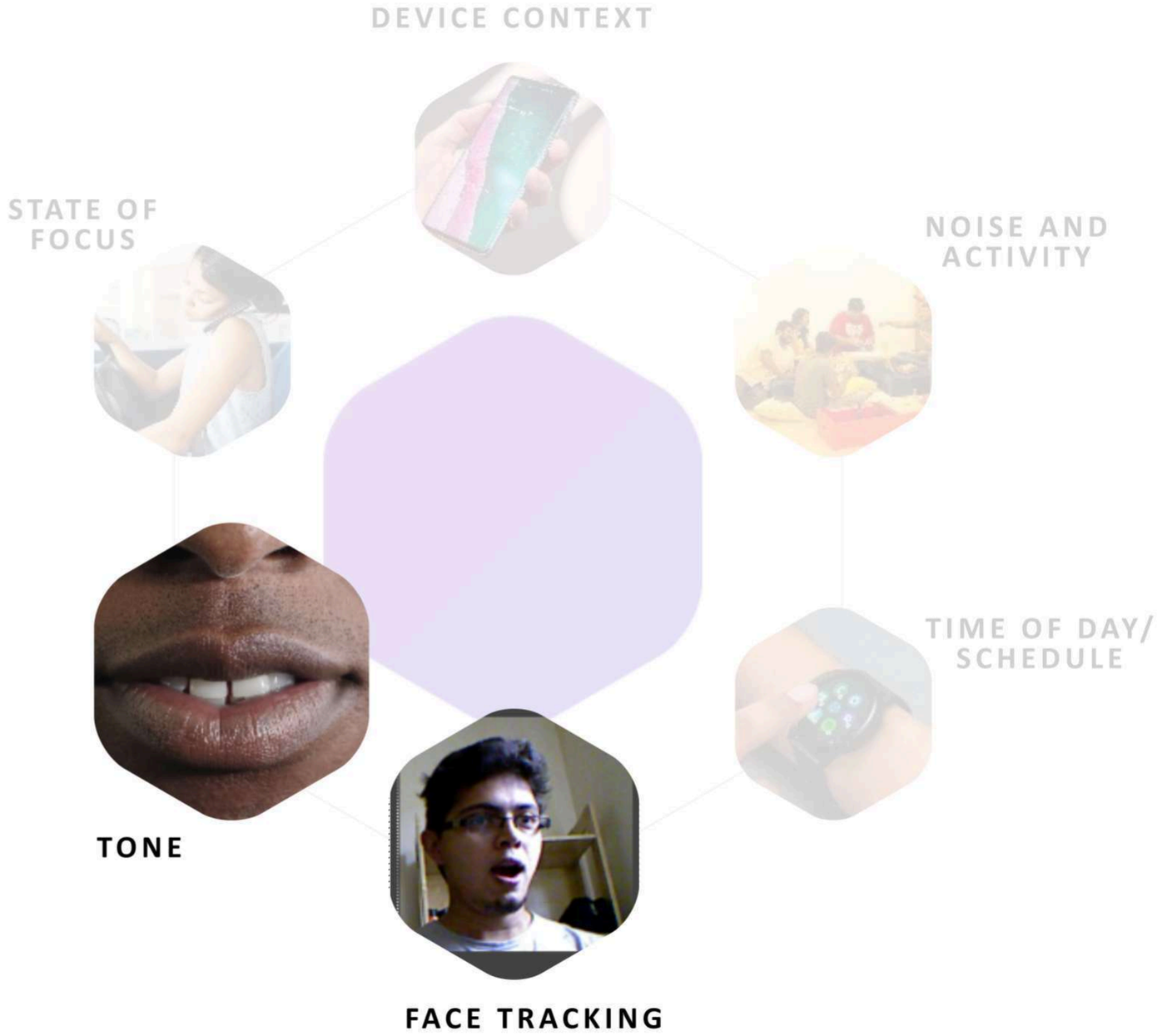
Agent recognizes the optimum time to notify the user



User gets some bad news

Agent recognizes the user's emotional state and assumes the appropriate tone

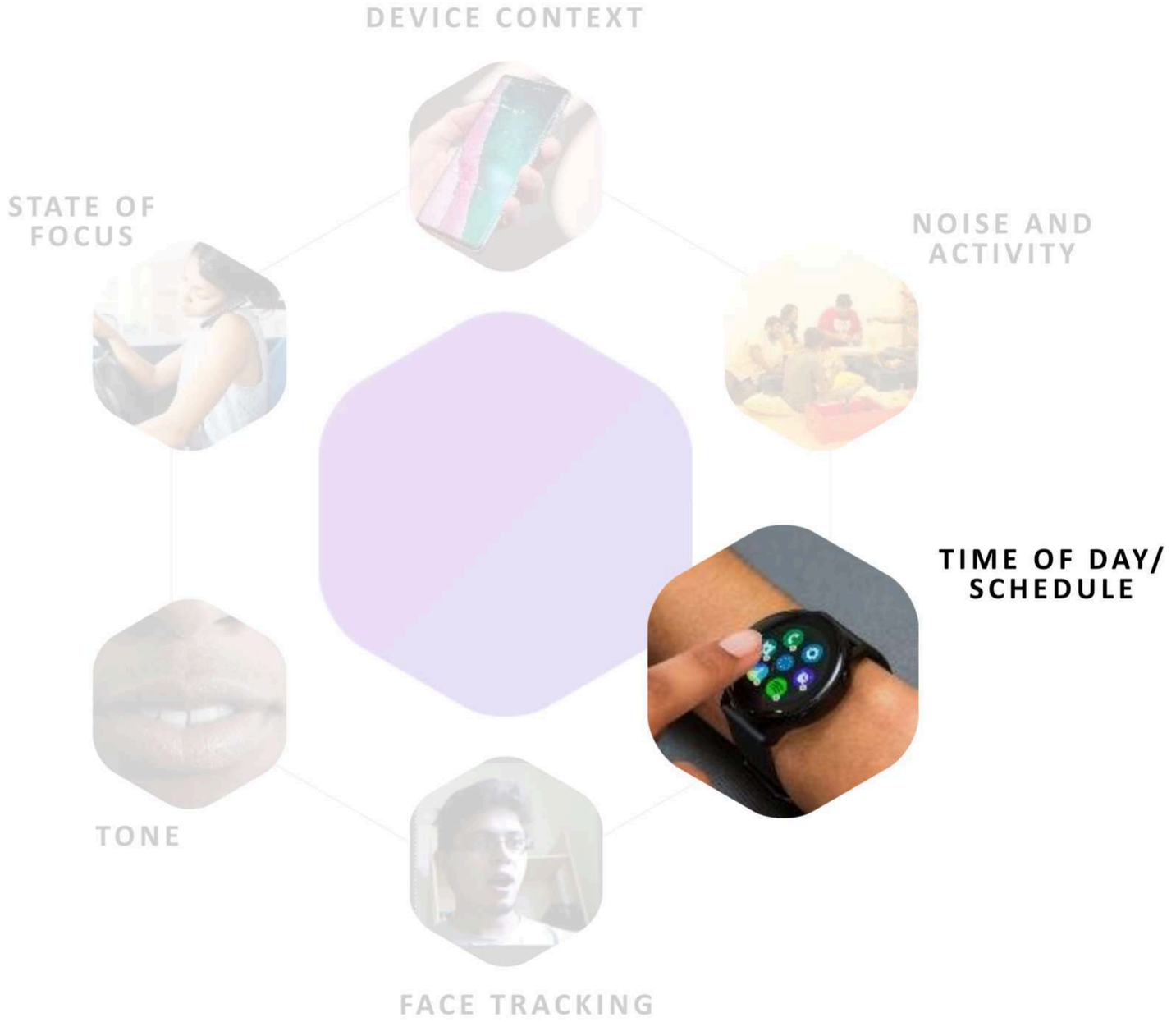
Agent acts sensitive to the situation



User is in a meeting

Agent recognizes this based on schedule and diverts calls

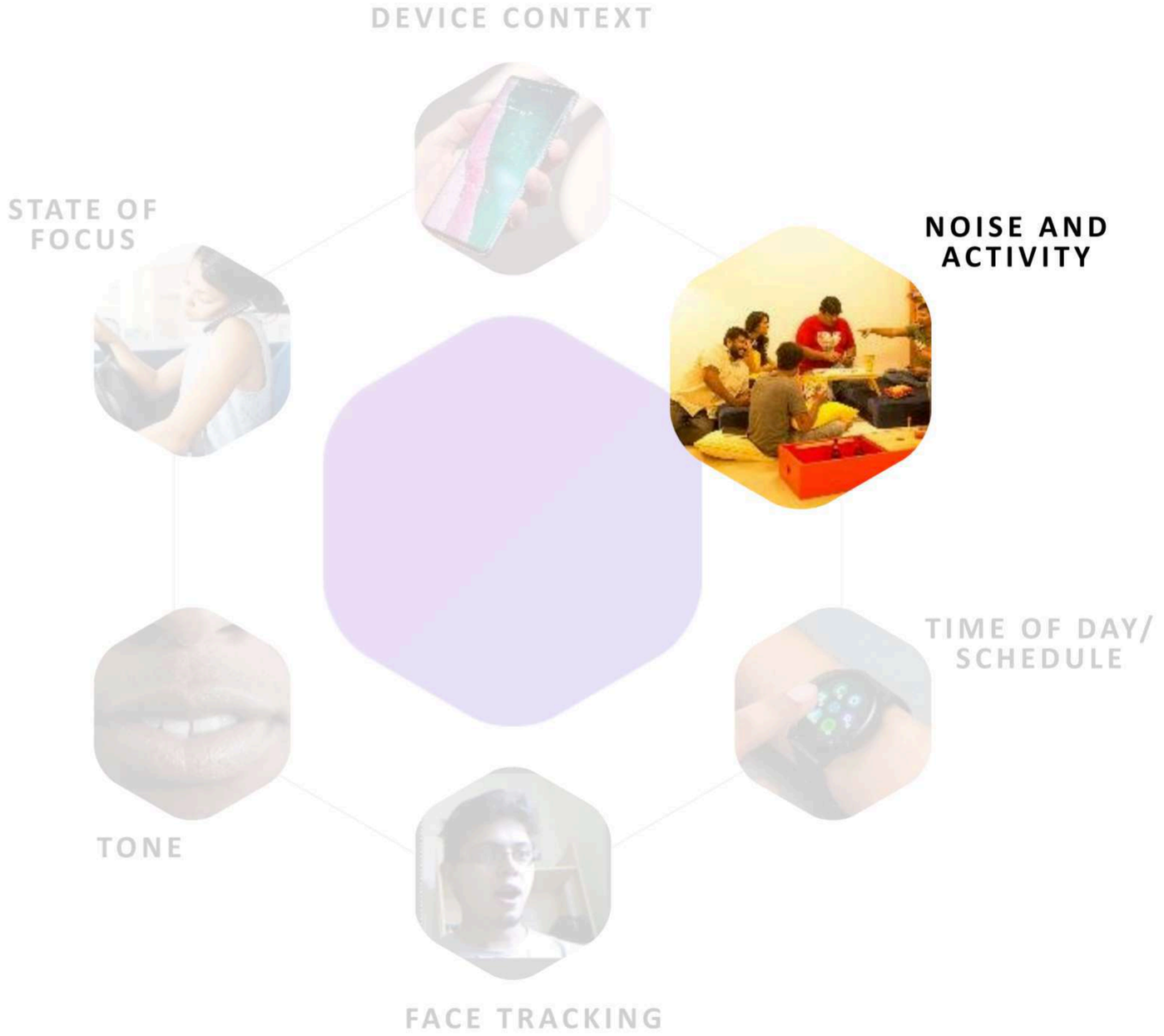
Agent alerts user on missed calls from important contacts



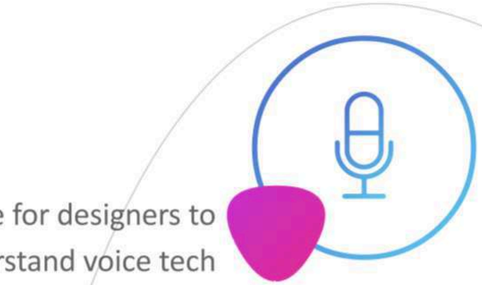
User has friends over

User has earlier told Agent to read emails at 8PM everyday

Agent recognizes context and doesn't do so



What is more natural?



UPDATE: A Blog/ Guide for designers to understand voice tech

HOW THE AGENT SOUNDS/ UNDERSTANDS YOU

Looking at ASR and TTS and understanding how they work.

INVOCATION

Looking at wake words and the different kinds of ways to initiate dialogue and sustain conversation.

CONTEXT

Figuring out how the agent can be more aware of the context user is in.

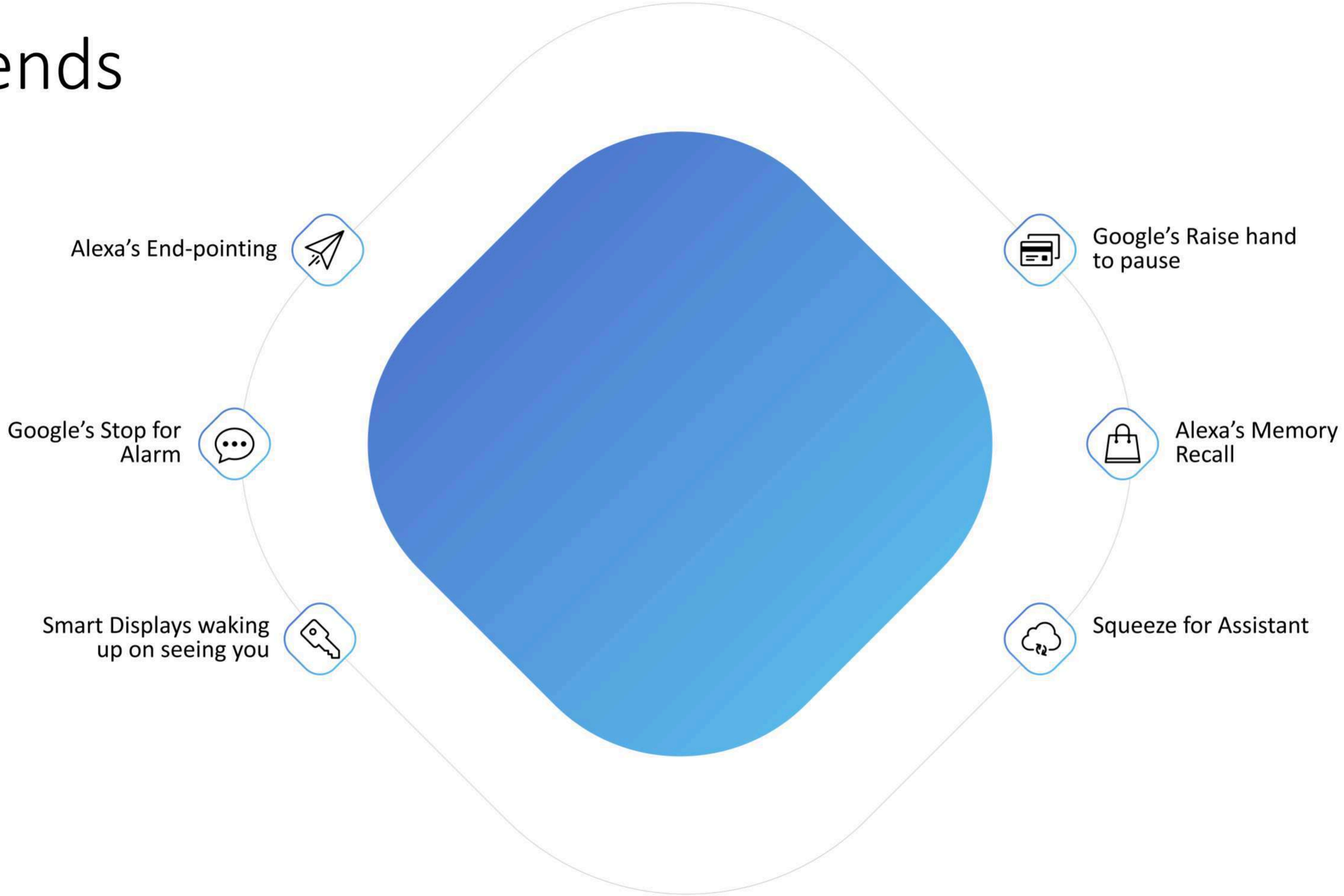
UPDATE: A Context aware system's design

Today’s Conversational Assistants

- Alexa, what’s the weather in Seattle on Sunday?
- Hey Cortana, open Netflix.
- OK Google, how tall is Mount Everest?

Barely a conversation.

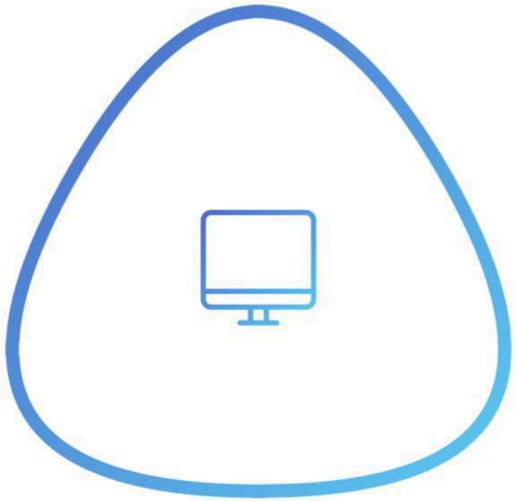
Trends


Alexa’s End-pointing
Google’s Stop for Alarm
Smart Displays waking up on seeing you
Google’s Raise hand to pause
Alexa’s Memory Recall
Squeeze for Assistant

In the near future we will have always listening devices

UX Process

FEEL IN CONTROL



TALK NATURALLY

MAKE TASKS EASIER

Concept

Talking to an Always-On Agent

Conversation Design Guidelines for such an agent

Design Principles

01

Maximize Efficiency

Allow non-wake word related invocations

02

Accommodate Errors

Foresee new kinds of errors

03

Maintain Transparency

If user asks, tell them why certain decisions were made

04

Ensure Clarity

User should be on the same page at all times

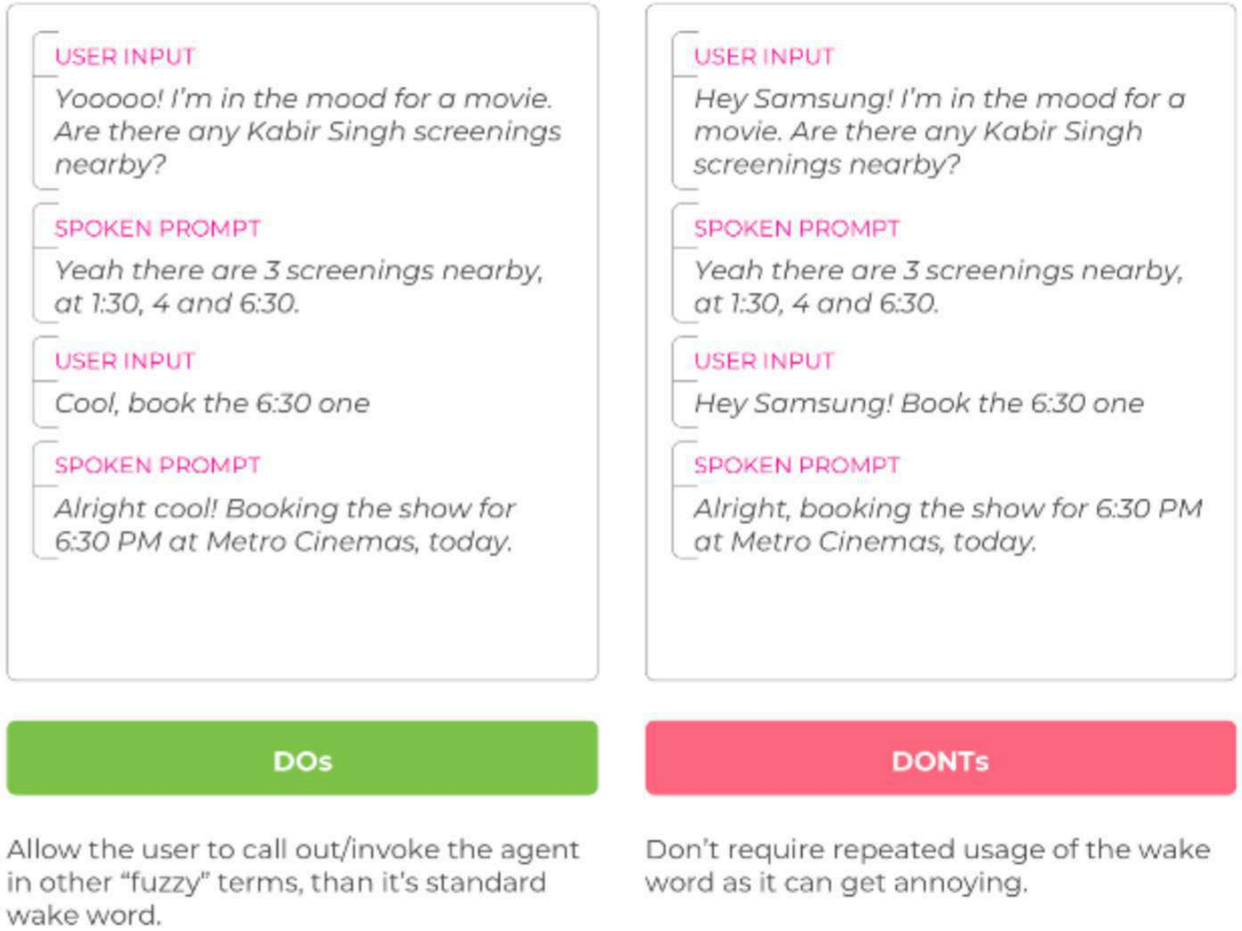
05

Provide Control

Make the user feel comfortable and in-control

01

Maximize Efficiency

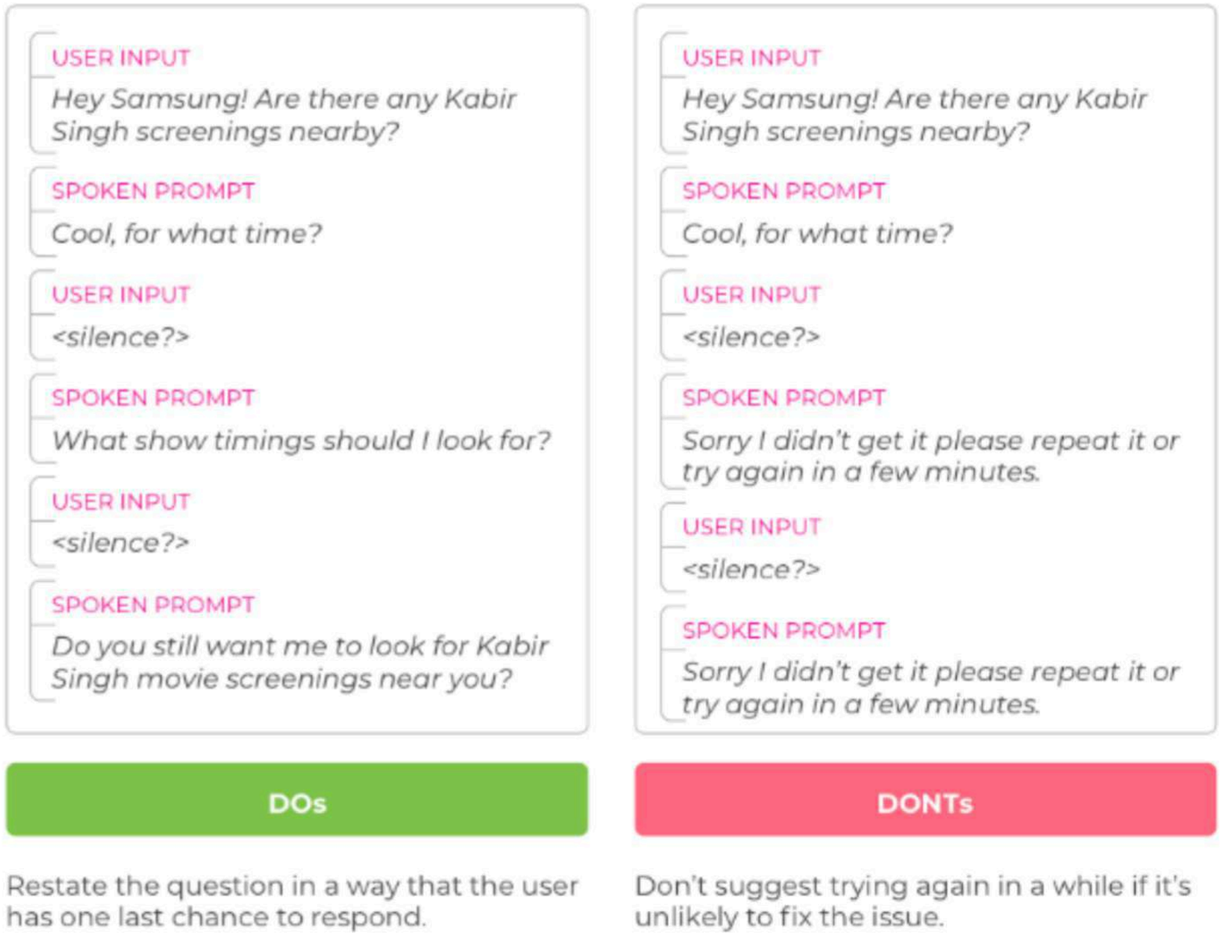


Allow the user to call out/invoke the agent in other "fuzzy" terms, than it's standard wake word.

Don't require repeated usage of the wake word as it can get annoying.

02

Accommodate Errors



03

Maintain Transparency

USER INPUT

Ummm... Samsung, why did the lights go out?

SPOKEN PROMPT

Kabir used the Philips Hue app to set a timer for these lights yesterday. The timer turns your downstairs lights off at 9PM every day.

DOs

Explain to the user clearly why some event happened and what app used to execute it so if they want to change things they can easily do so.

USER INPUT

Ummm... Samsung, why did the lights go out?

SPOKEN PROMPT

I'm not sure about that.

SPOKEN PROMPT

Philips Hue's legal team doesn't allow me to. disclose that information.

DONTs

Not giving the users a clear answer may make them feel uncomfortable, scared and out of control of their own devices.
Putting the blame on others can also backfire, offer them the answer and take responsibility.

04

Ensure Clarity

USER INPUT

Hey, I added that number for Jenny the other day right?

SPOKEN PROMPT

Yep! You added +91 987654321

USER INPUT

Send this to Saahil.

SPOKEN PROMPT

Sending Jenny's new number to Saahil through text.

DOs

Reiterating the points acknowledge that the agent heard the user correctly.

USER INPUT

Hey, I added that number for Jenny the other day right?

SPOKEN PROMPT

Yep!

USER INPUT

Send this to Saahil.

SPOKEN PROMPT

Done.

DONTs

Without the reiteration the user might feel unsure about if the agent has heard them correctly.

05

Provide Control

BARGE IN

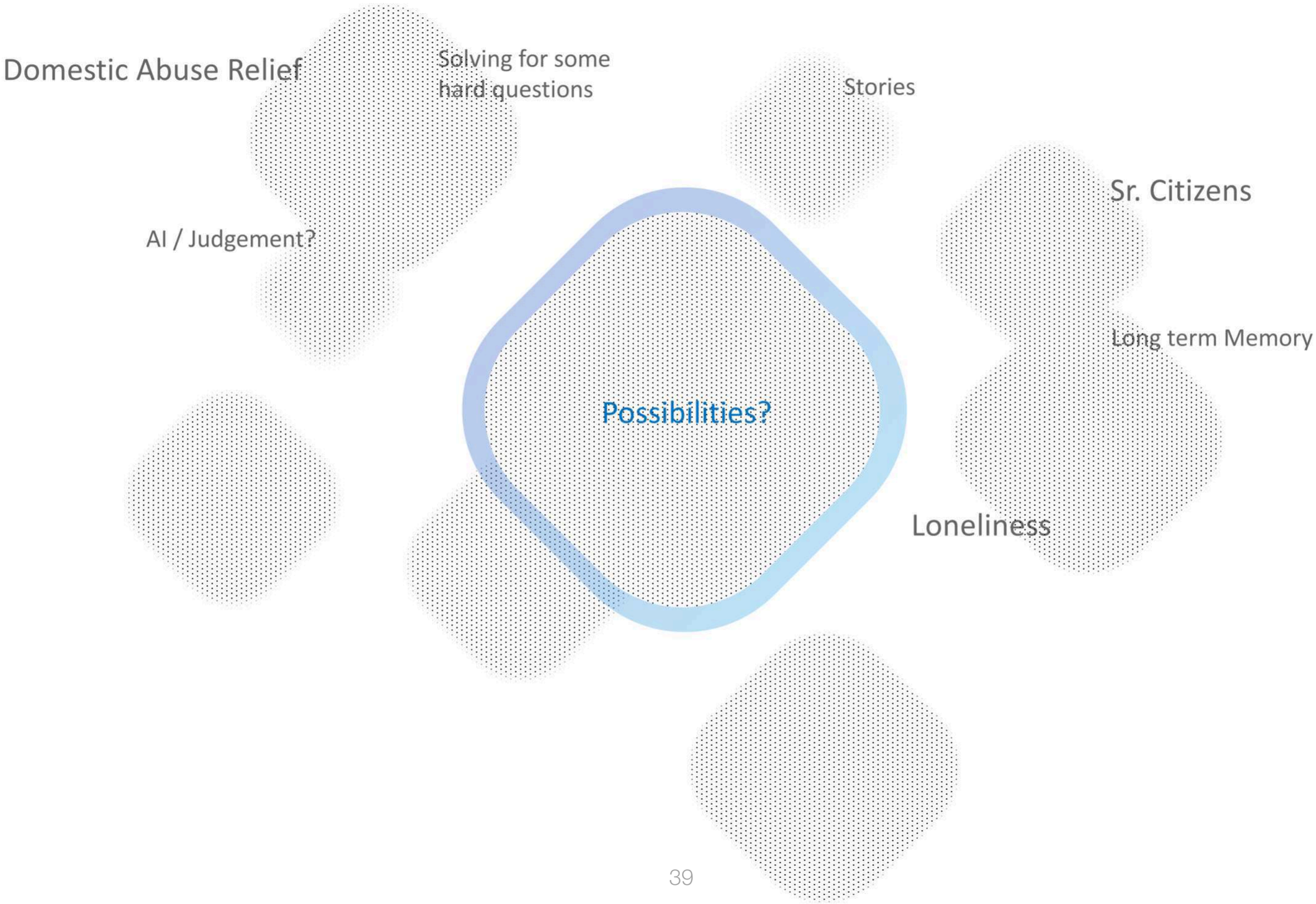
Allow the user to start speaking while the agent still is going on and on

RECOVERY

Allow the user to feel undo any command they had given, easily

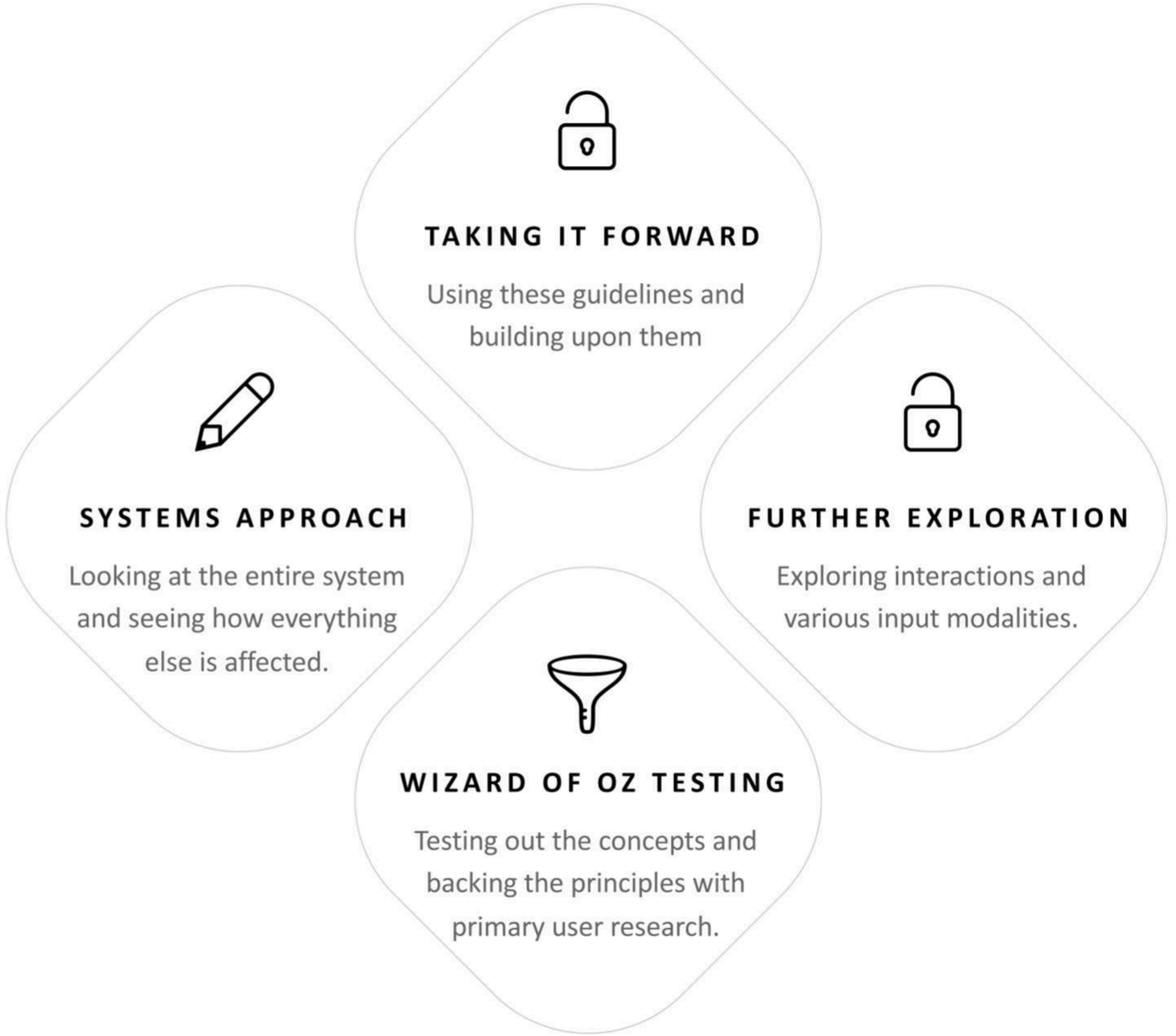
CUT OFF

Make it easy for the user to make the agent stop listening


Domestic Abuse Relief
Solving for some hard questions
Stories
Sr. Citizens
AI / Judgement?
Long term Memory
Possibilities?
Loneliness

Future Scope

What’s next for the project?



Key Learnings

EMBRACE AMBIGUITY

When exploring such a new field, almost nothing has been defined. And that's okay.

MISTAKES

It's alright to make mistakes as long as you learn from them and move ahead.

HOW LARGE COMPANIES WORK

How shipping some feature or a product happens at such a scale

MAKE TIME FOR FUN AS WELL

Loving your work is fine, but if you overwork yourself… You'll get burnt out.

BE FLEXIBLE

Move fast, break things but learn from your mistakes and the advice of others as well!

OPEN ENDED PROJECTS

Doing such an open ended project from scratch… Is really hard

Thank You

LOOKING FORWARD TO YOUR FEEDBACK