



Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής

Ενότητα 6.2: Ανάπτυξη Διαλογικών
Φωνητικών Εφαρμογών
Γεώργιος Κουρουπέτρογλου

koupe@di.uoa.gr



Περιεχόμενα ενότητας

Μεθοδολογίες και εργαλεία ανάπτυξης φωνητικών
διαλογικών εφαρμογών

Διαχείριση Διαλόγων: Πρόσφατες Τάσεις

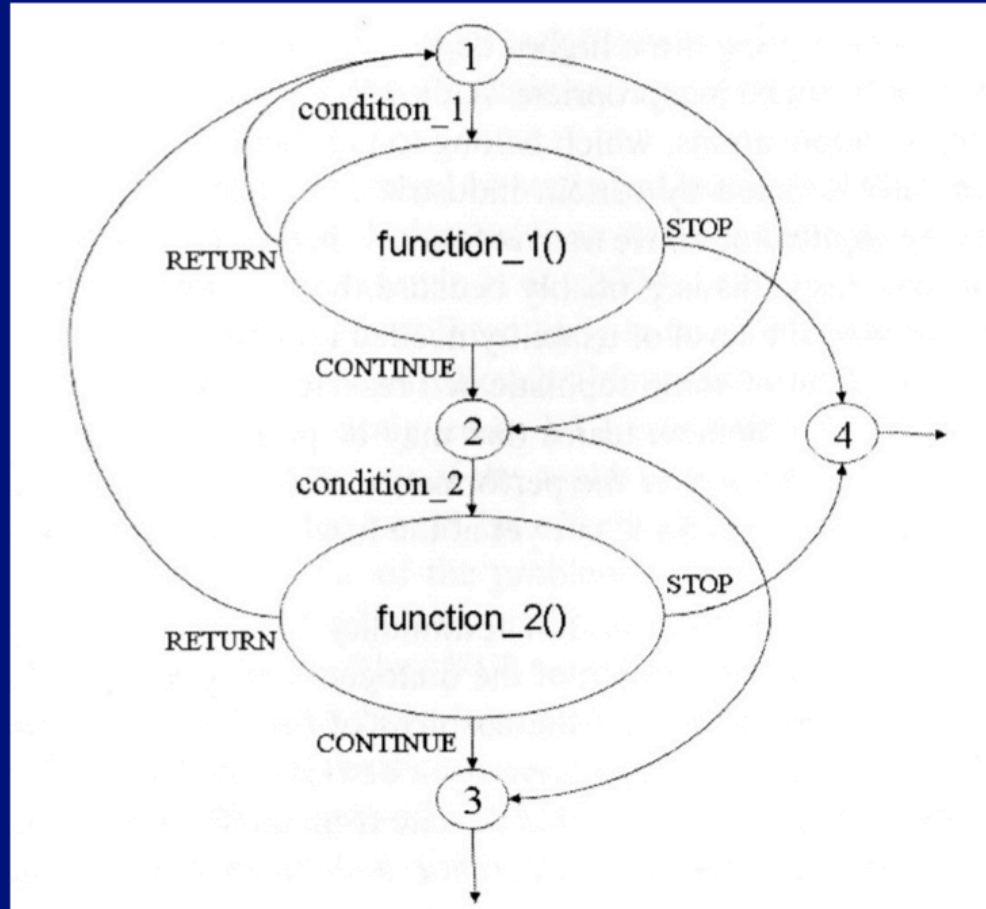
Πρόσφατες Τάσεις

- Fixed Topology Models
 - Μικρές ειδικευμένες γραμματικές
 - Frame-like structure (κατασκευή)
- Task-based dialog management using an agenda
- Plan-based dialogue
 - Εξαγωγή πλάνου διαλόγου
 - Νέα εξαγωγή/επαλήθευση/διόρθωση/αναδιοργάνωση
- Inference-based Dialogue Manager
 - Problem solving εφαρμογή

Fixed Topology (1 από 2)

- Καθορισμός παραμέτρων μικρής εμβέλειας
- Μικρές ειδικευμένες γραμματικές
- Frame-like structure
- Σταθερή και διακριτή συμβολή κάθε (υπο) γράφου
- Flags για μετάβαση στην τοπολογία

Fixed Topology (2 από 2)



Task-based dialog management using an agenda (1 από 5)

A1: Hello. You've reached the [Communicator]. Tell me your full name

U2: Hi I'd like to fly to seattle tuesday morning

A3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?

U4: [John Doe]

A5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.

A6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?

U7: Yes

A8: Will you return to Pittsburgh from Seattle?



Task-based dialog management using an agenda (2 από 5)

U9: no i will go to Los Angeles

A10: Traveling to Los Angeles. on what date

U11: Sorry can we go back to talk about the previous leg again

A12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?

U13: actually i want to leave on wednesday

A14: a flight on wednesday.

A15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?



Task-based dialog management using an agenda (3 από 5)

U16: Yes

A17: OK and on what day do you want to depart
Seattle?

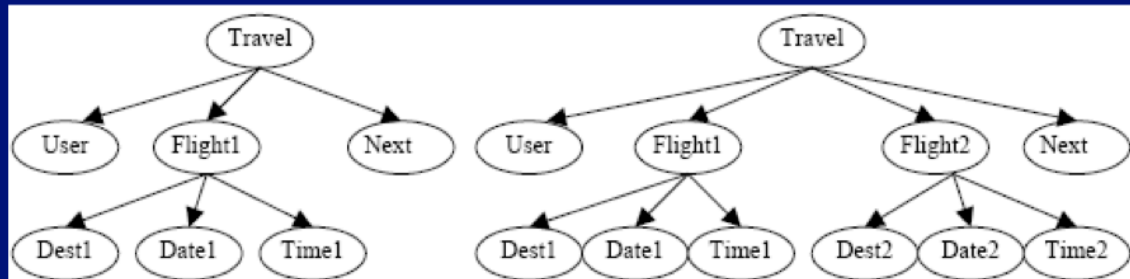
U18: next Monday

A19: Traveling on Monday, August sixteenth.

Task-based dialog management using an agenda (4 από 5)

- Προσδοκίες
- Κανόνες
 - Αναγνώρισης
 - Ενημέρωσης
 - Επιλογής
 - Κατανόησης
- Agenda με μερικώς ταξινομημένους στόχους (goals)

Task-based dialog management using an agenda (5 από 5)



The initial product tree (simplified)

Product tree at utterance A10

Travel	User	User	Flight1	Next	Date2	Flight1	Flight1	Date2
User	Travel	Travel	Dest1	Flight1	Flight2	Dest1	Dest1	Flight2
Flight1	Flight1	Flight1	Date1	Dest1	Dest2	Dest1	Date1	Dest2
Dest1	Dest1	Dest1	Time1	Date1	Time2	Time1	Time1	Time2
Date1	Date1	Date1	User	Time1	Next	Date2	Date2	Flight1
Time1	Time1	Time1	Travel	User	Flight1	Flight2	Flight2	Dest1
Next	Next	Next	Next	Travel	Dest1	Dest2	Dest2	Date1
					Date1	Time2	Time2	Time1
					Time1	Next	Next	Next
					User	User	User	User
					Travel	Travel	Travel	Travel

initial

A1

A3

A6

A8

A10

A12

A15

A17

Plan-Based Dialog Management Structure (1 από 5)

- Μηχανή reasoning (συλλογισμού)
- Στρατηγική remembering – forgetting
- Η στρατηγική δεν είναι δεμένη σε θεματικό πεδίο (domain-independent)
- Το μοντέλο διαχείρισης διαλόγου εύκολα μεταφέρεται μεταξύ θεματικών περιοχών

Plan-Based Dialog Management Structure (2 από 5)

- ΠΛΑΝΟ

- Παρακολουθεί και επιτηρεί το συνολικό διάλογο
- Παρακολουθεί το ποσοστό της συμπλήρωσης του στόχου (task).
- Υπολογίζει τα επόμενα βήματα για τη συμπλήρωση του στόχου
- Συντηρεί την αδιάλειπτη πρόοδο προς τη συμπλήρωση του στόχου.
- Θέτει ενδιάμεσους στόχους για την υποβοήθηση της επιτυχούς έκβασης του τελικού στόχου

Plan-Based Dialog Management Structure (3 από 5)

- ΕΥΡΩΣΤΙΑ

- Χειρίζεται τις παρεκκλίσεις από την υπολογισμένη πρόοδο προς λύση, π.χ. ο χρήστης μπορεί να ρωτήσει κάτι το οποίο είναι άσχετο ή όχι σωστό
- Προτείνει και επιβάλλει αμοιβαία ασύμβατους περιορισμούς για να μην παρεκκλίνει ο διάλογος
 - π.χ. ο χρήστης μπορεί να κάνει λάθος ερώτηση ή το σύστημα να μην καταλάβει και ο διάλογος να ξεφύγει του στόχου.
 - π.χ. ο χρήστης μπορεί να μην ορίσει επαρκώς το αίτημα και το σύστημα να δώσει άμεσα μια τελική λύση χωρίς αυτή τελικά να είναι αποδεκτή από το χρήστη

Plan-Based Dialog Management Structure (4 από 5)

- Αρχικός καθορισμός πλάνου
- Νέα εξαγωγή
- Επαλήθευση
- Διόρθωση
- Αναδιοργάνωση

Plan-Based Dialog Management Structure (5 από 5)

U1: I would fly to New York.	(User initiative, telling the arrival city)
S1: What day are you going to fly there?	(System initiative, asking for departure time)
U2: Wednesday.	(User answering)
S2: There are several flights. When will you prefer to depart?	(System initiative, asking for departure time)
U3: After 10:00.	(User answering)
S3: Which flight do you prefer, 10:10 or 16:30?	(System initiative, asking for departure time)
U4: What's the arrival time for the flight at 16:30?	(User initiative, asking for arrival time)
S4: 19:10.	(System answering)
U5: What's the local time in Beijing then?	(User initiative, topic changed, asking for time difference)
S5: 08:10.	(System answering)
U6: Fine, I want the flight departing at 16:30.	(User initiative, topic changed, verifying departure time)
S6: You want Flight CA 981 on 14 th June, departing from Beijing at 16:30 and arriving in New York at 19:10, is that correct?	(System initiative, asking for user confirmation)
U7: Yes, exactly.	(User confirming)

Inference-based dialogue manager

- Problem solving
- Υψηλή πολυπλοκότητα -> Υψηλό σχεδιασμό (planning)
- Αποφεύγουν τον συσχετισμό καταστάσεων (states) και πράξεων (acts)
- Ανάλυση της θεματικής περιοχής σε Goals και subgoals
 - Δενδροειδής αναπαράσταση των goals/subgoals
 - Ιεραρχία αποφάσεων βασισμένη σε οντολογία στόχων
 - Δυναμική διάρθρωση δομής βασισμένη σε agenda -> λίστα subgoals (subgoal = συλλογή πληροφοριών)
- Με στατιστικά μοντέλα, μηχανική μάθηση -> μαθαίνει στρατηγική

Dialogue acts (1 από 3)

forward acts	assert, info-req, order, request, thank, greeting, closing, express, check, suggest, promise, offer, apology, encourage, accuse, intro-topic, avoid
backward acts	accept, reject, address, answer, divert, counterpropose, hold, clarify-parameter, redirect, confirm
conversation	start-conversation, end-conversation, confirm-start, deny-start, pre-close
grounding	initiate, continue, repair, acknowledge, request-repair, cancel
turn-taking	keep-turn, hold-turn, release-turn, assign-turn
initiative	take-initiative

Dialogue acts (2 από 3)

- Core speech acts (forward, backward)
- Dialogue acts (everything else)

Dialogue acts (3 από 3)

ΕΠΕΝΕΡΓΟΥΝ ΚΑΙ ΔΕΣΜΕΥΟΥΝ ΤΟΥΣ ΕΤΑΙΡΟΥΣ ΤΟΥ ΔΙΑΛΟΓΟΥ

- Assertions (ισχυρισμοί)
 - Δέσμευση του εκφωνούντα ότι η πράξη έγινε/γίνεται/θα γίνει
- Info-requests (αιτήματα πληροφοριών)
 - Ερώτηση που απαιτεί απάντηση ή λύση
- Requests (αιτήματα για πράξη)
 - Απαίτηση λύσης
- Order (διαταγή)
 - Απαίτηση δέσμευσης από ανώτερο σε κατώτερο για πράξη
- Suggestions (εισήγηση)
 - Δεν υπάρχει δέσμευση παρά μόνο εστίαση στο θέμα

Διαλογικά Συστήματα: Χρησιμότητα - Αξιολόγηση χρησιμότητας

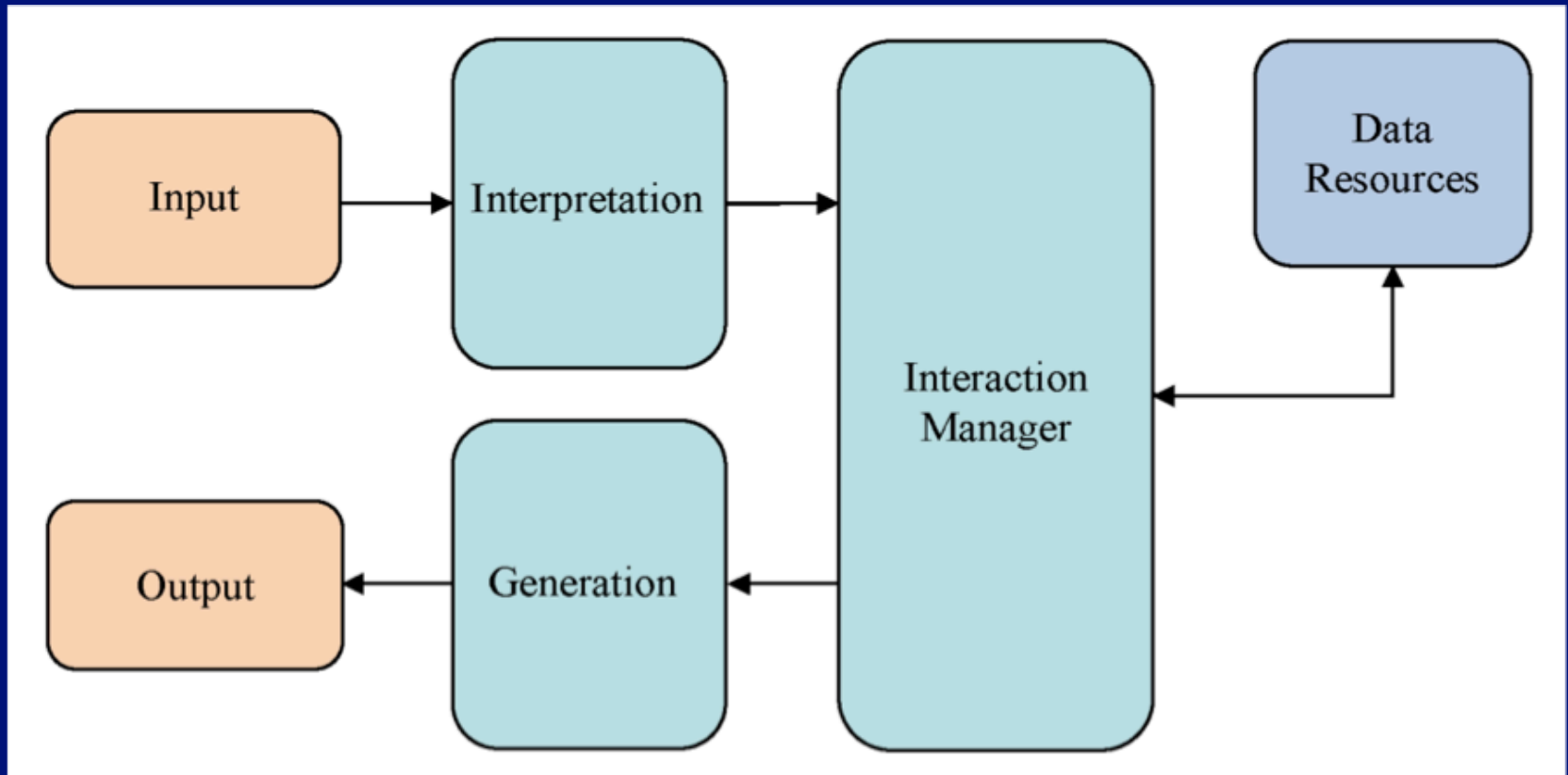
Χρησιμότητα (1 από 4)

- Usefulness (χρησιμότητα)
 - Κατά πόσο ο χρήστης θα ήθελε να το χρησιμοποιήσει (task enablement)
- Efficiency (αποδοτικότητα)
 - Ταχύτητα – ακρίβεια – αρτιότητα - πληρότητα
- Effectiveness (αποτελεσματικότητα)
 - Συμπεριφορά συστήματος σε σχέση με το πώς ο χρήστης περιμένει ότι θα συμπεριφερθεί υο σύστημα
 - Επίπεδο προσπάθειας του χρήστη για να φτάσει σε επιθυμητούς στόχους (goals)
 - Δυσκολία επίτευξης στόχων χρήστη

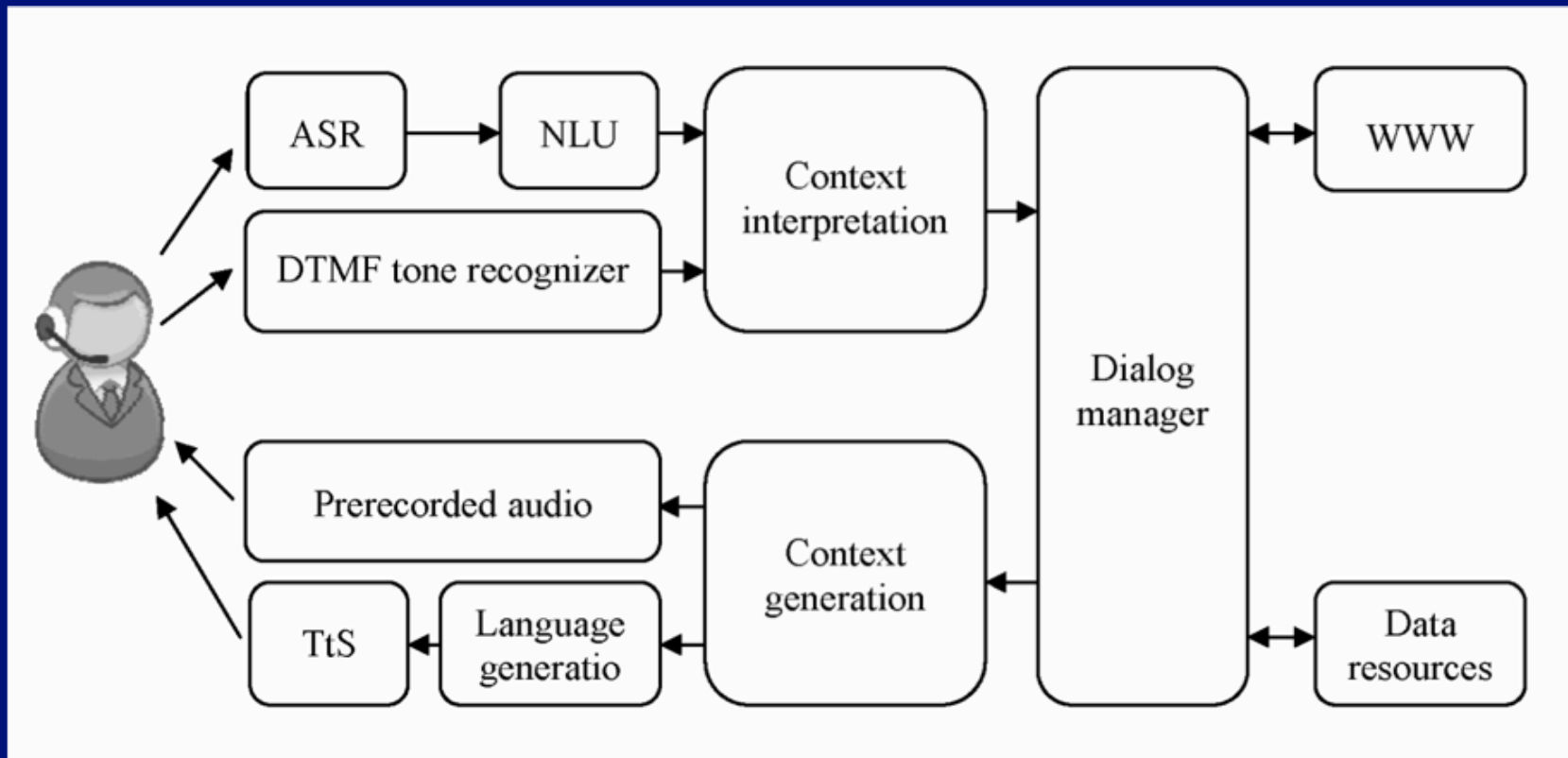
Χρησιτικότητα (2 από 4)

- Learnability (ικανότητα εκμάθησης)
 - Επεκτείνει την αποτελεσματικότητα
 - Μετράει τη δοκιμασία , το βίωμα του χρήστη (User experience)
 - Μετράει την ελάφρυνση της δυσκολίας με το χρόνο και τις επαναλήψεις συγκεκριμένων στόχων
- Satisfaction
 - Μετράει την συνολική εικόνα και υποκειμενική γνώμη του χρήστη
- (Accessibility)

Χρησιμότητα (3 από 4)



Χρησιμότητα (4 από 4)



Κύκλος ζωής (lifecycle) (1 από 3)

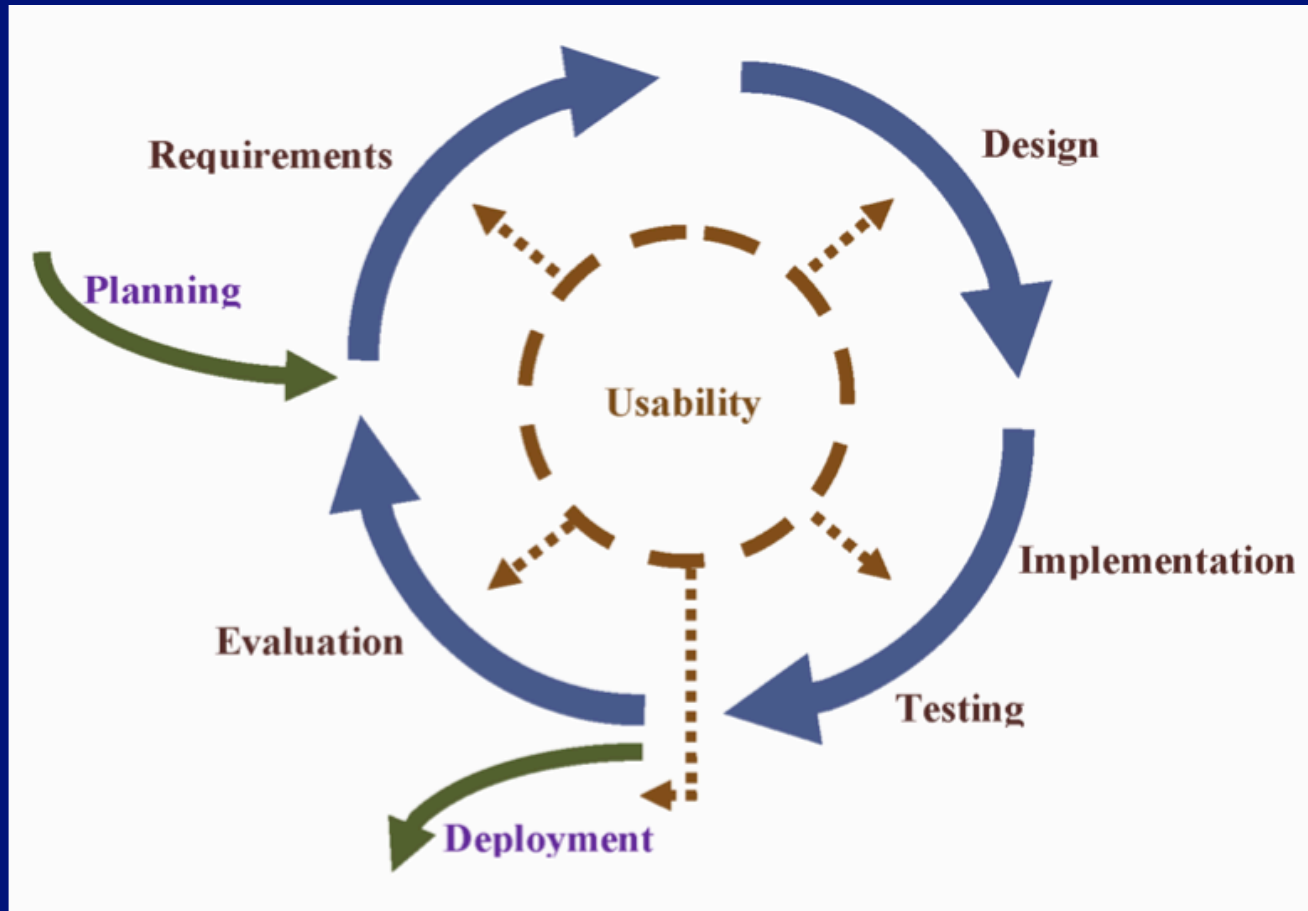
There are five activities in the lifecycle of a speech interface (Sharp et al., 2007):

- Requirements specification and initial planning
- Design
- Implementation and testing
- Deployment
- Evaluation

Κύκλος ζωής (lifecycle) (2 από 3)

1. User involvement should take place throughout all five stages.
2. The usability requirements, goals and evaluation parameters should be set at the start of the development.
3. Iteration through the five stages is inevitable and, therefore, should be included in the initial planning.

Κύκλος ζωής (lifecycle) (3 από 3)



Τύποι Διαλογικών Συστημάτων και Χρηστικότητα (1 από 2)

Τα διαλογικά συστήματα είναι συνήθως 3 τύπων όσον αφορά το σχεδιασμό τους:

(a) DTMF replacement

(b) Simple system or user-directed question-answering

(c) Open-ended natural language mixed-initiative conversational system.

Τύποι Διαλογικών Συστημάτων και Χρησιμότητα (2 από 2)

Αντίκτυπος του σχεδιασμού στη χρησιμότητα με
βάση τον κύκλο ζωής:

Τύπος	Requirements	Design	Implementation	Deployment	Evaluation
a	low	medium	low	low	low
b	medium	medium	low	low	medium
c	high	high	medium	high	high

Τυπικές απαιτήσεις (1 από 3)

- ΑΠΑΙΤΗΣΕΙΣ

- Μη-λειτουργικές - αφορούν τα κριτήρια λειτουργίας του συστήματος
- Λειτουργικές - αφορούν τη συμπεριφορά (behaviour) του συστήματος

1. User satisfaction – Users should be satisfied or very satisfied either as standalone users or comparing their input from using an earlier interface.
2. Quality of service offering – Improvement on the quality of the way the requested services/tasks are presented. For example, a large DTMF tree-based dialogue may require the users to navigate through several menu layers to achieve their goal, while a natural language dialogue may identify the initial user request and retrieve the requested service right at the start.
3. State-of-the art solution – The system should deploy cutting edge technology.

Τυπικές απαιτήσεις (2 από 3)

4. Ability to provide customised behavioural or personalised interaction for specific user groups. A common example is the use of a preferred type of interaction (formal, casual, friendly, entertaining, etc.) set specifically for the application domain.
5. Complete access to all services or business units that are supported. By design, the system should be able to provide the users the same high quality interaction for all services that the interface is used for.
6. Reliability – Extends to the system providing the intended functions continuously without failing.
7. Continuity of processing – Also includes problem recovery. In this case a natural language interface should cater for the interaction when a system problem occurs.

Τυπικές απαιτήσεις (3 από 3)

8. Auditability – Ensures the transparency of the system providing supporting evidence to trace processing of data.
9. Performance requirements that describe the capacity of the system to perform certain functions or process certain volume of transactions within a prescribed time.
10. Usability-related factors that the operator of a spoken dialogue interface may find prudent to stress upon to the designer.

Acceptance Tests (1 από 3)

- Key performance indicators (KPIs) for ASR and TtS success
- Functional assessment of the system respective modules and functions such as accuracy of information relayed to the user, start/end of dialogue or sub-dialogue flow, service/information provision accuracy, and so on.
- User Experience assessment in terms of:
 - a. quality issues
 - b. user interaction

Acceptance Tests (2 από 3)

- User Experience assessment in terms of:
 - a. Quality issues
 - speech or dialogue pause length between activities such as voice request, system search, information retrieval, information relay/output, and prompt delays between responses
 - output voice (natural or synthetic) consistency and naturalness for all stages of dialogue as well as in special cases where critical information or explicit help is required
 - choice of presenting output voice, clear and non-breaking, during loudspeaker mode or in noisy environments
 - correct pronunciation and focus placement in sentences

Acceptance Tests (3 από 3)

b. User interaction

- ease of use, navigation through the interface
- instructions and help prompt quality
- smart recovery from misinterpretations or misrecognitions
- disambiguation and confirmation function performance
- dialogue flow cohesion
- overall satisfaction.

Μεθοδολογίες (1 από 9)

The Wizard-of-Oz formative evaluation:

- Early testing – it can be performed in the early stages in order to test and formulate the design parameters as early in the product lifecycle as possible.
- Use of prototype or early design – eliminates problems arising later in the development such as integration.
- Language resources – Grammar coverage for the speech recognition (ASR) and respective machine learning approaches for interpretation (NLU) are always low when testing a non-finalised product. Low scoring for ASR-NLU may hinder the usability evaluation, however, the use of the human usability expert eliminates such handicap.
- System updates – the system, being a mock-up, can be updated effortlessly to accommodate for changes imposed from the input from the test subjects, making it easier to re-test the updated system in the next usability evaluation session.

Μεθοδολογίες (2 από 9)

The results of the WOZ tests are both from the user subjective feedback and the examination of the objective performance measures.

The performance measures include:

- task completion – whether the participants completed the specified tasks that were set within the scenarios successfully,
- efficiency – whether the participants chose the most direct route to the goal, using the predetermined scenario feedback to compare against the optimal path for the same scenario that was expected by the designer,
- dialogue flow – how the participants chose to interact with the system, the number of times the help was requested and how informative it was, as well as the number of times disambiguation, confirmation and error recovery sub-dialogues were enabled.

Μεθοδολογίες (3 από 9)

The subjective input of the participants is recorded through questionnaires that the participants fill in after each task completion as well as at the end of the evaluation.

The questions are used to assess the user experience asking about complexity, effort required, efficiency, linguistic clarity, simplicity, predictability, accuracy, suitable tempo, consistency, precision, forgiveness, responsiveness, appropriateness, overall impression and acceptance of the system, either regarding particular tasks or the full system.

- Μετριοούνται σε Likert scales (1-to-5 or 1-to-7 scales)

(for example, 1 being “totally disagree” and 7 “totally agree” and the rest in between)

Μεθοδολογίες (4 από 9)

- The summative usability testing:

At this stage, the usability tests play a much more pivotal role since the development of the system is near completion. There are three distinct purposes for usability testing of a working system: the development, testing and tuning. During the development the users test a nearly finished product, during testing a finished product, and during tuning a finished and already deployed product. Regardless of purpose, the tests focus on all aspects that the WOZ handled as well as several aspects that the WOZ ignored:

- Grammar testing
- Interpretation testing
- Dialogue management/flow
- System response adequacy
- Output speech quality.

Μεθοδολογίες (5 από 9)

For spoken dialogue interfaces, the following 15 objective (both quantitative and qualitative) and subjective usability evaluation criteria have been proposed (Dybkjær & Bernsen 2000):

1. Modality appropriateness.
2. Input recognition adequacy.
3. Naturalness of user speech relative to the task(s) including coverage of user vocabulary and grammar.
4. Output voice quality.
5. Output phrasing adequacy.
6. Feedback adequacy.

Μεθοδολογίες (6 από 9)

7. Adequacy of dialogue initiative relative to the task(s).
8. Naturalness of the dialogue structure relative to the task(s).
9. Sufficiency of task and domain coverage.
10. Sufficiency of the system's reasoning capabilities.
11. Sufficiency of interaction guidance (information about system capabilities, limitations and operations).
12. Error handling adequacy.

Μεθοδολογίες (7 από 9)

13. Sufficiency of adaptation to user differences.

14. Number of interaction problems (Bernsen et al. 1998).

15. User satisfaction.

Μεθοδολογίες (8 από 9)

Apart from task completion and dialogue flow, depending on the domain, as a general rule, functional measurements should be recorded for at least the following indicative parameters:

- Average call duration
- Peaks and valleys of usage per hour per day
- Successful speech recognitions
- Misrecognitions
- No-inputs
- Timeouts

Μεθοδολογίες (9 από 9)

- Rejections
- Early hang-ups
- Successful interpretations
- Failed interpretations (no-matches)
- Successful repairs
- Failed repairs.

Τέλος Ενότητας

Ανάπτυξη Διαλογικών Φωνητικών Εφαρμογών

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «**Εκπαίδευση και Δια Βίου Μάθηση**» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

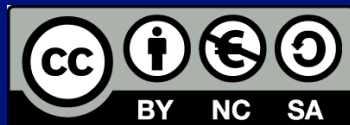
Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Γεώργιος Κουρουπέτρογλου 2015. Γεώργιος Κουρουπέτρογλου.
«Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής. Ανάπτυξη Διαλογικών Φωνητικών Εφαρμογών». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/DI37/>.

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων (1 από 2)

Οι φωτογραφίες που περιέχονται στην παρουσίαση αποτελούν πνευματική ιδιοκτησία τρίτων. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση και διάθεσή τους στο κοινό με οποιονδήποτε τρόπο χωρίς τη λήψη άδειας από τους δικαιούχους.

Σημείωμα Χρήσης Έργων Τρίτων (2 από 2)

Η δομή και οργάνωση της παρουσίασης, καθώς και το υπόλοιπο περιεχόμενο, αποτελούν πνευματική ιδιοκτησία του συγγραφέα και του Πανεπιστημίου Αθηνών και διατίθενται με άδεια Creative Commons Αναφορά Μη Εμπορική Χρήση Παρόμοια Διανομή Έκδοση 4.0 ή μεταγενέστερη.