Speech To Text Alignment
oooo

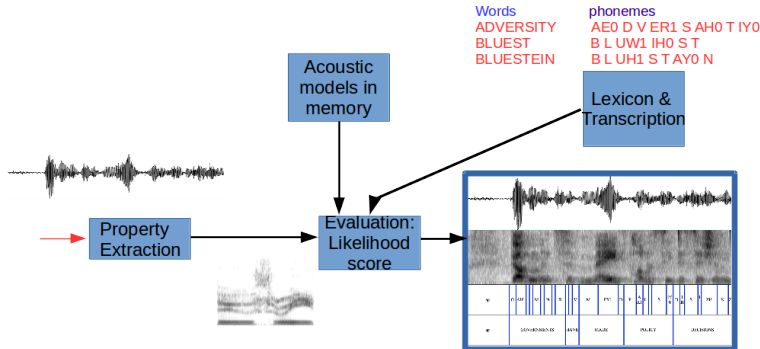Prosody
ooo

Summary
ooooo

# Enhance Speech Text Alignment For Prosody Modeling And Prediction In TTS

Aghilas Sini

Université de Rennes 1

13/01/2017

Speech To Text Alignment
● ○ ○ ○

Prosody
○ ○ ○

Summary
○ ○ ○ ○ ○

Overview

# Forced Alignment

Speech To Text Alignment
○●○○

Prosody
○○○

Summary
○○○○○

Adapting The Default Acoustic Model

# Why?

- Adapt data properties (words, phonemes ... )
- Voice characteristics.
- Take account recording environment (robustness)
- Accurate boundaries (words, silences ... )

Speech To Text Alignment
○○●○

Prosody
○○○

Summary
○○○○○

Problems In Text Processing

# Text Pre-processing(Classical Problems)

Since the texts are too long

- sentence end detection (choice should be done..)
- Dealing with abbreviations.
- Recognizing Acronyms and URLs.
- Processing numbers
- Dealing with idioms and proper name.

Note : I'm trying to solve the above problems using different existing toolkit, this introduce another issues... And I think that we need manual checking ...

# Gaëlle Vidal Work to Complete!!

| CORPUS NEB (monolocuteur) | DUREE | Durée sélection Qualité | words/sec, base txt | % st.dir base txt | Durée fichiers annot. man. Expressivité | Durée fichiers annot. Man. personnages | Durée lab. Man st. Dir | Dmoy(sec) lab. man. st. Dir | % st.dir base annot. man | Durée ΔTimbre | % D ΔTimbre /st.dir |
|---|---|---|---|---|---|---|---|---|---|---|---|
| mysteres | 25:31:34 | 25:09:09 | 2,68 | 68% | 07:33:17 | 07:33:17 | 04:39:20 | 7,9 | 62% | 04:24:30 | 58% |
| miserables | 14:16:15 | 10:58:50 | 2,46 | 16% | | 14:10:22 | 02:06:07 | 5,1 | 15% | 00:51:35 | 6% |
| mme Bovary | 13:11:41 | 12:45:32 | 2,36 | 20% | 00:41:48 | 00:41:48 | 00:01:48 | 4,5 | 4% | 00:00:27 | 1% |
| vampire (la) | 10:02:22 | 07:35:55 | 2,47 | 37% | 00:05:45 | 00:05:45 | 00:00:05 | 2,7 | 1% | | 0% |
| fille du pirate | 06:29:32 | 02:09:35 | 2,52 | 30% | 02:08:47 | 02:08:47 | 00:25:27 | 5,2 | 20% | 00:08:38 | 7% |
| roman de la momie | 06:32:33 | 06:32:33 | 2,40 | 24% | 00:24:00 | 00:24:00 | 00:03:44 | 5,3 | 16% | 00:01:49 | 8% |
| carmen | 02:17:50 | 01:24:20 | 2,67 | 23% | 01:24:02 | 01:24:02 | 00:17:04 | 4,9 | 20% | 00:06:08 | 7% |
| voyage centre terre | 01:51:38 | 01:29:20 | 2,43 | 19% | | | | | | | |
| geminal | 01:45:50 | 00:55:30 | 2,58 | 14% | | 01:45:35 | 00:12:03 | 4,7 | 11% | 00:03:32 | 3% |
| contes cruels | 01:42:47 | 01:17:59 | 2,47 | 32% | | 01:41:45 | 00:26:19 | 6,7 | 26% | 00:00:48 | 1% |
| contes | 01:05:47 | 01:05:47 | 2,77 | 27% | 00:10:50 | 01:05:23 | 00:21:13 | 5,4 | 32% | 00:12:14 | 19% |
| venus | 01:01:44 | 01:01:44 | 2,93 | 43% | 01:01:32 | 01:01:32 | 00:25:12 | 7,2 | 41% | 00:07:36 | 12% |
| lettres persanes | 00:20:27 | 00:20:27 | 2,61 | 89% | | 00:20:01 | 00:16:49 | 144,2 | 84% | 00:06:26 | 32% |
| fables la fontaine | 00:18:06 | 00:13:00 | 2,53 | 12% | | 00:17:37 | 00:04:01 | 6,2 | 23% | 00:00:46 | 4% |
| chants de maldoror | 00:17:37 | 00:10:38 | 2,35 | 0% | | 00:17:28 | 00:00:08 | 2,7 | 1% | 00:00:08 | 1% |
| poemes | 00:11:52 | 00:11:16 | 2,21 | 2% | | 00:11:17 | 00:06:47 | 27,1 | 60% | | 0% |
| infernaliana | 00:10:34 | 00:10:34 | 2,67 | 35% | | 00:10:26 | 00:04:14 | 7,3 | 41% | 00:00:29 | 5% |
| cerfvolant | 00:05:49 | 00:00:00 | 2,41 | 6% | | 00:05:44 | 00:02:18 | 2,9 | 5% | 00:00:18 | 5% |
| TOTAL | 87:13:58 | 73:32:09 | 2,53 | 38% | 13:30:01 | 33:24:49 | 09:15:11 | | 28% | 06:03:43 | 18% |

Remarques

= durée des fichiers son

Voir Document Description du corpus

= Estimations à partir du texte

= calculs à partir des étiquetages manuels.
Les durée totales peuvent légèrement différer des durées de signal du fait de l'absence de labels sur les silences de fin de fichiers.

(italique = lectures à la 1ère personne)

Prosody Modeling

# Some Related Work

Speech To Text Alignment

0000

Prosody

○●○

Summary

00000

Prosody Prediction

# Some Related Work

# Questions?

Speech To Text Alignment
OOOO

Prosody
OOO

Summary
●OOOO

Future scope of work

# Suggested problems to work on

- direct and indirect discourse.

Speech To Text Alignment
○○○○

Prosody
○○○

**Summary**
○●○○○

Work Plan(Schudele)

# Planning the unplannable?

Speech To Text Alignment

oooo

Prosody

ooo

Summary

oo●oo

Work Plan(Schudele)
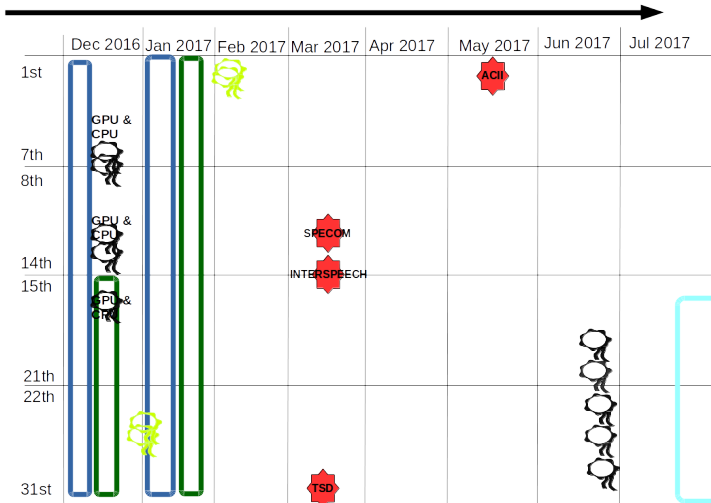
THANKYOU

# References

- Simon, Anne-Catherine; Auchlin, Antoine; Avanzi, Mathieu & Goldman, Jean-Philippe. Les phonostyles: une description prosodique des styles de parole en français, Rev. Abecassi, M. & G. Ledegen, Les voix des Français. En parlant, en écrivant, Peter Lang ,**2010**, 71-88
- Sandrine, Brognaux; Thomas, Drugman & Marco, Saerens. Synthesizing Sports Commentaries: One or Several Emphatic Stresses? Rev. Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), Social and Linguistic Speech Prosody. Proceedings of the 7th international conference on Speech Prosody, **2014**, 270-274
- Brognaux, Sandrine; Picart, Benjamin & Drugman, Thomas. Synthèse de commentaires sportifs: intégration d'une annotation prosodique à deux niveaux dans un synthétiseur HMM. Rev. Nouveaux Cahiers de la Linguistique Française. **2014**, 31, 335.

Speech To Text Alignment      Prosody      Summary
OOOO      OOO      OOOOO●
References

# References (Continued)

- Urbica, T.; Vlacy, V.; Kalyuzhnyi, Y. V.; Dill, K. A. An Imporoved Thermodynamic Peterbation Theory for Mercedes-Benz Water J. Chem. Phy. **2007**, 127, 1-4

- Silverstein, K. A. T.; Haymet, A. J. D.; Dill, K. A. Molecular Model of Hydrophobic Solvation J. Chem. Phys. **1999**, 111(17), 8000-8009