

TELISMA



TELISMA, European leader in speech technologies



***How can the ASR research
community***

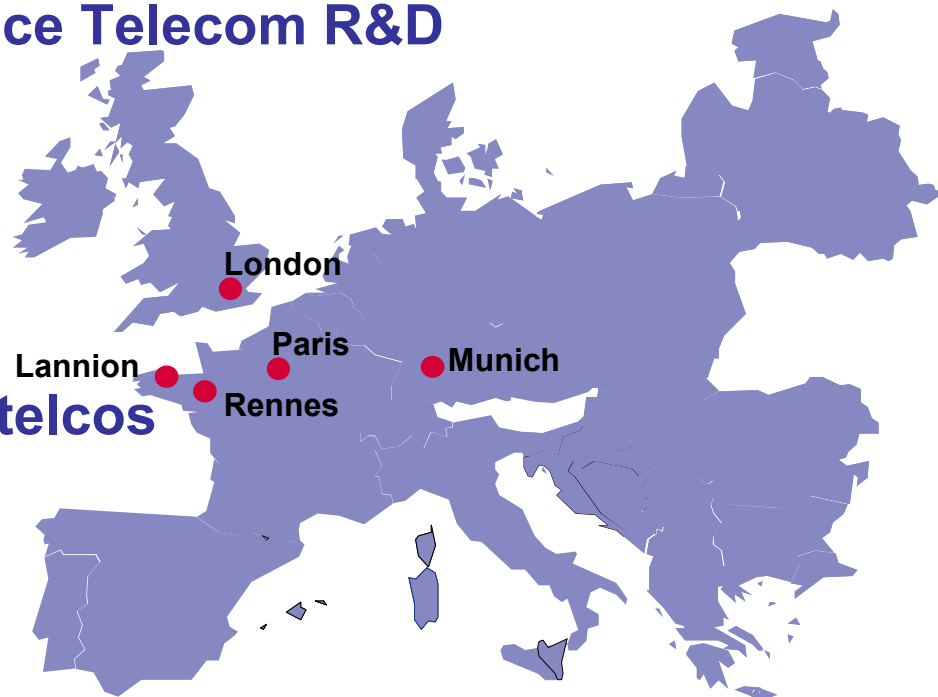
***Help the ASR industry
to deliver best products,
applications and services***

***Francis Charpentier, Chief Scientific Officer
fcharpentier@telisma.com***

telisma highlights



- Founded: **August 2000**
Spin-off from France Telecom R&D
- Headquarters: **Paris**
- Sales offices: **Munich**
London
- Headcount: **55 employees**
- Mission: **ASR technologies for telcos**



500 M calls will use telisma's technology in 2003
36 000 telephony ports already deployed in Europe

Are we going to face an ASR industrial slowdown ?



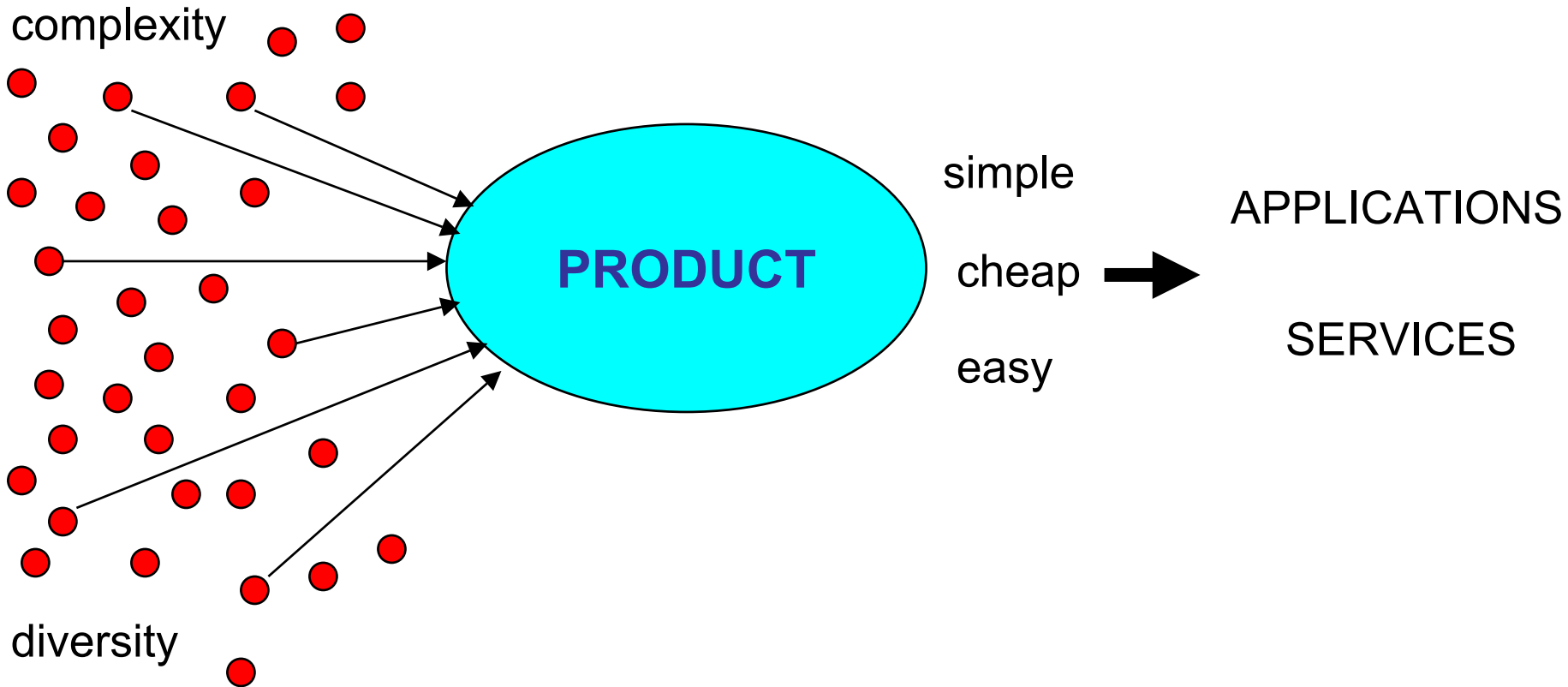
ASR industry is put through higher pressure than ever

- Deploying applications remains very expensive
- ASR industry continues to loose money
- Most ASR customers don't care
- We are in a buyers' market :
 - Customers take time to decide
 - They negotiate low prices
- But we are also in a dreamers' market :
 - Unrealistic expectations regarding performance
 - Risky assumptions regarding ease of app development (VoiceXML for all)
 - Customers believe ASR should be a commodity !

The ASR industry challenge



RESEARCH → ENGINEERING → MARKET



From worldwide ASR research to a worldwide ASR market



A large worldwide scientific community

A handful of international players

Equipment manufacturers

Telecom operators

Enterprises

ASR research community

ASR industry

ASR customers

Expertise

Creativity

Enthusiasm

Thriving to deliver quality products

Intensive customer support

High economic pressure

ASR not a priority

Should be off-the-shelf

Should be cheap

From worldwide ASR research to a worldwide ASR market

Speech is an exciting subject

Speech needs a long term determination

VUI is an exciting interface

**ASR research
community**

*Expertise
Creativity
Enthusiasm*

**ASR
industry**

*Thriving to deliver quality
products
Intensive customer support
High economic pressure*

ASR customers

*But ASR not a priority
Should be off-the-shelf
Should be cheap*

Feet on the ground !

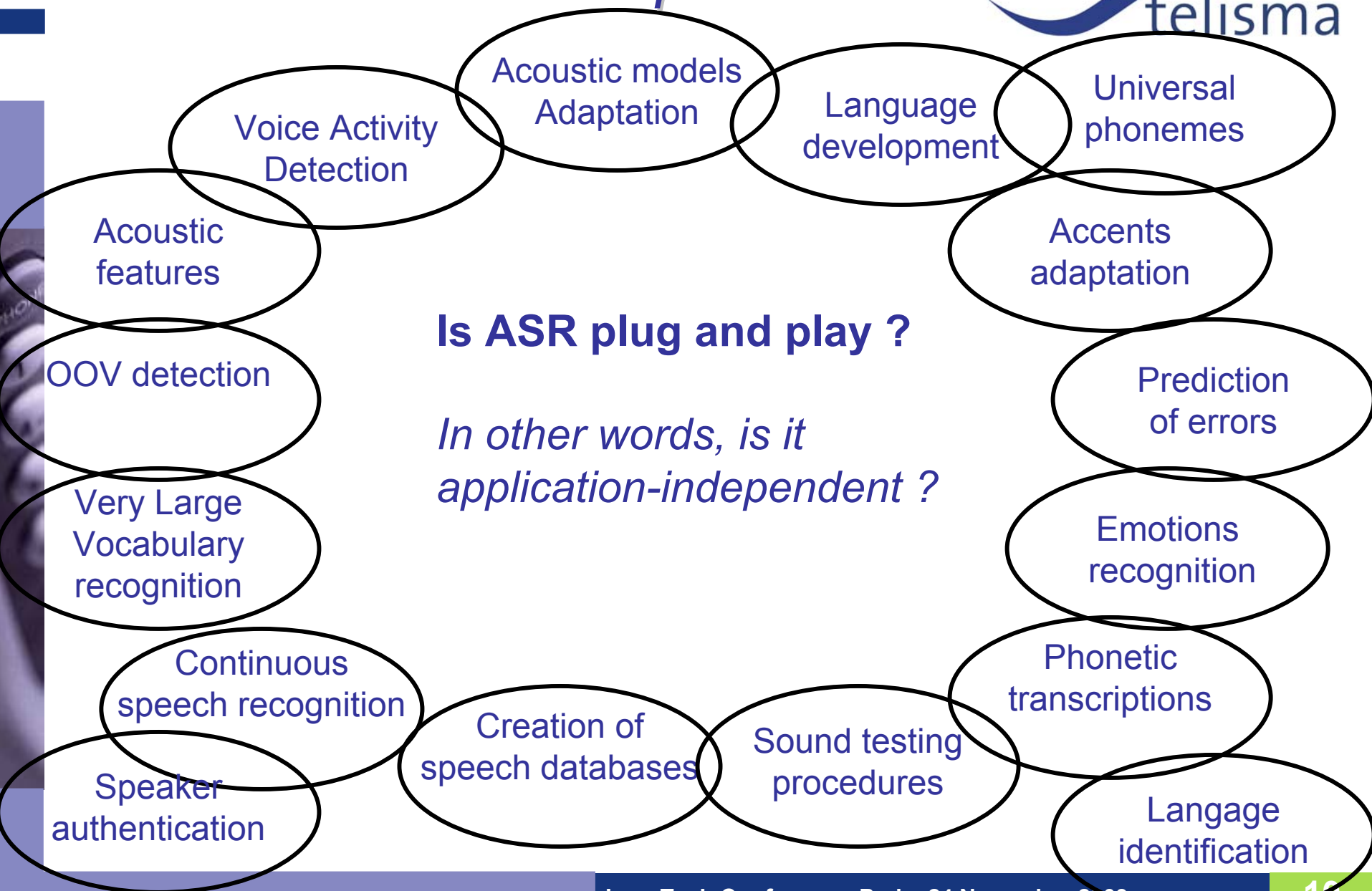


- Let's stop dreaming
- Speech works today, but it is expensive
- ASR is a long way before becoming a commodity
- To work, ASR applications require specific optimisation , specific optimisation, and even more specific optimisation...
- To succeed, ASR industry needs more cooperation both from the scientific community and from the market

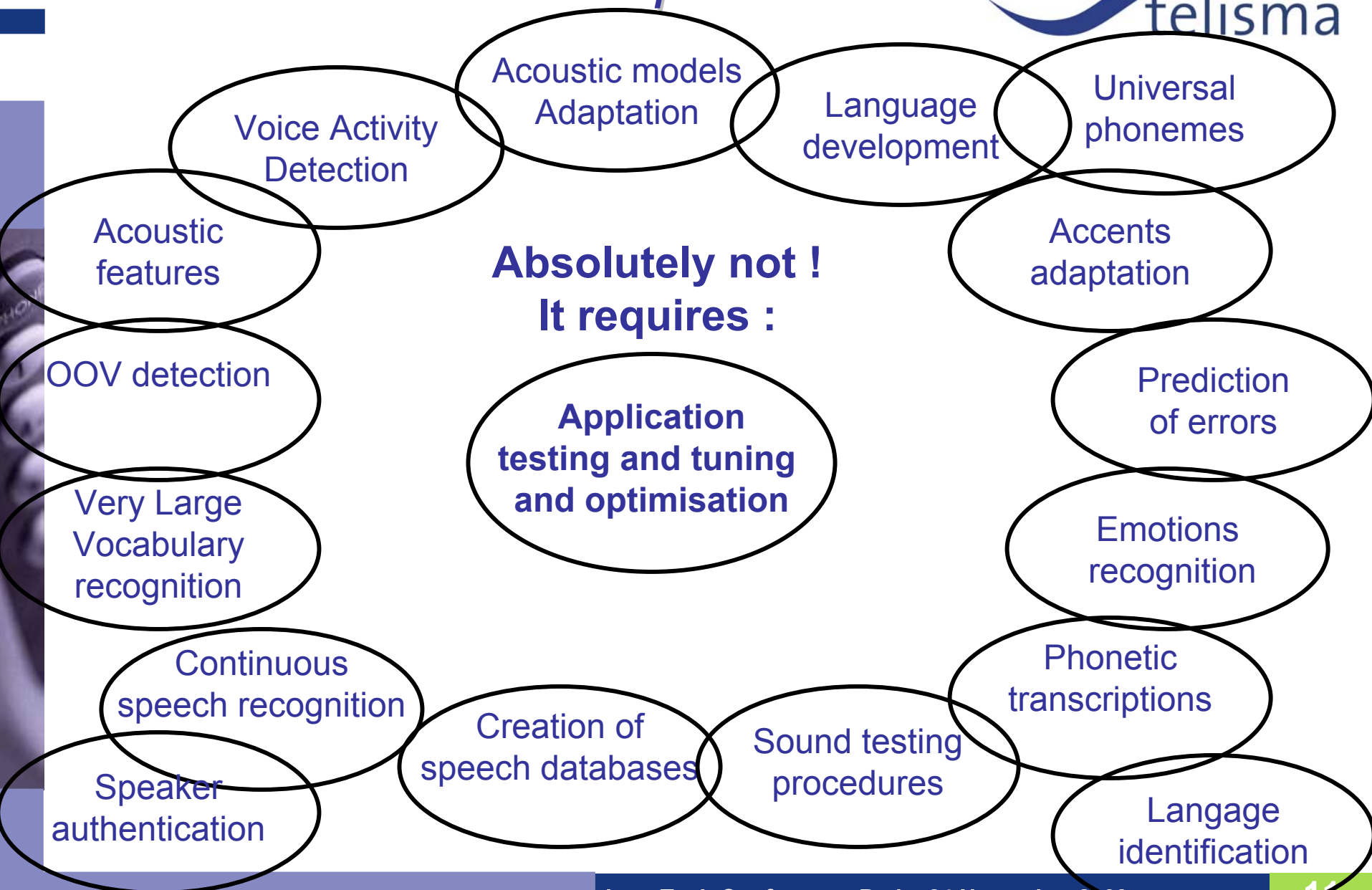
Overview of ASR components



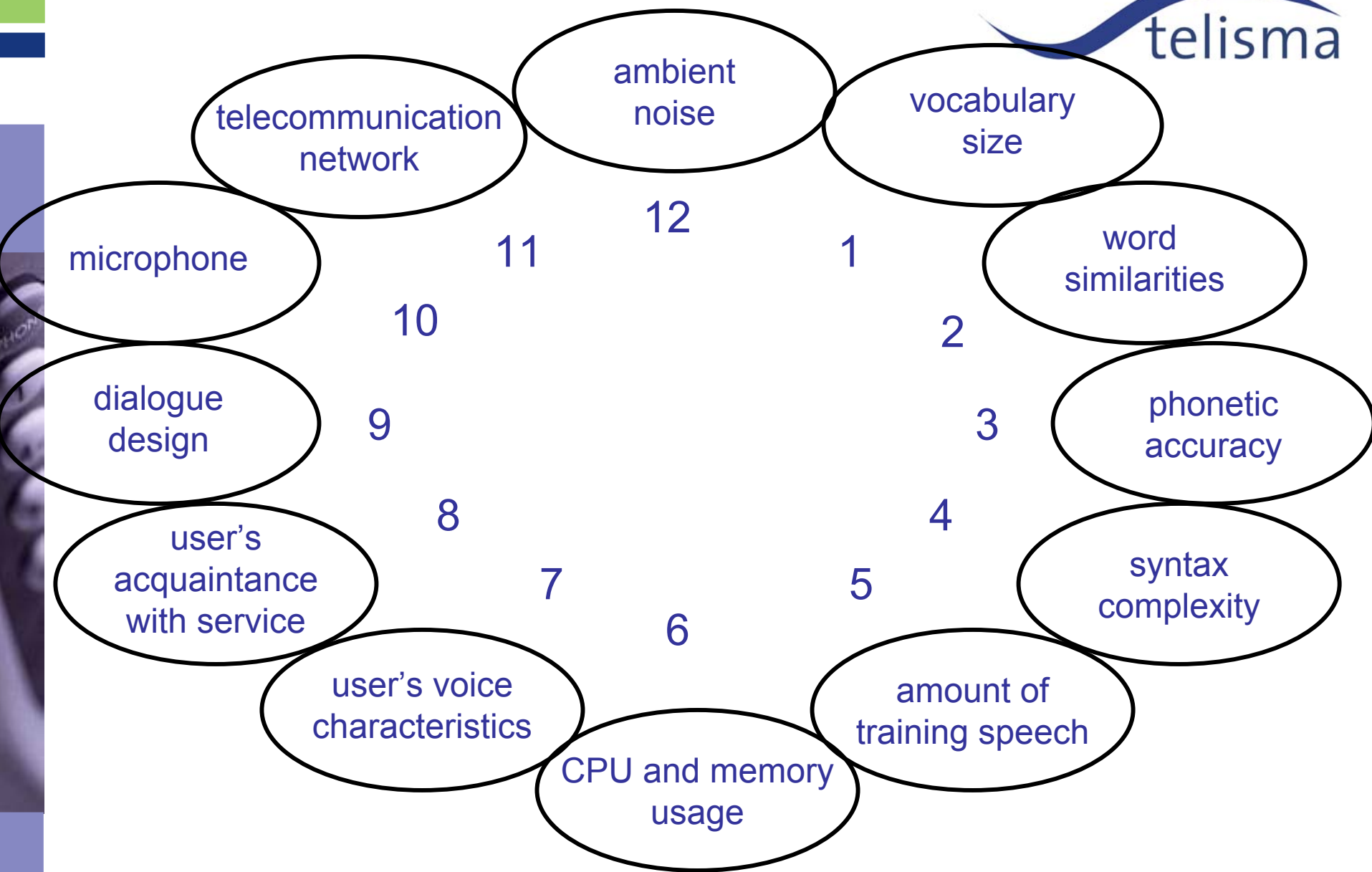
Overview of ASR components



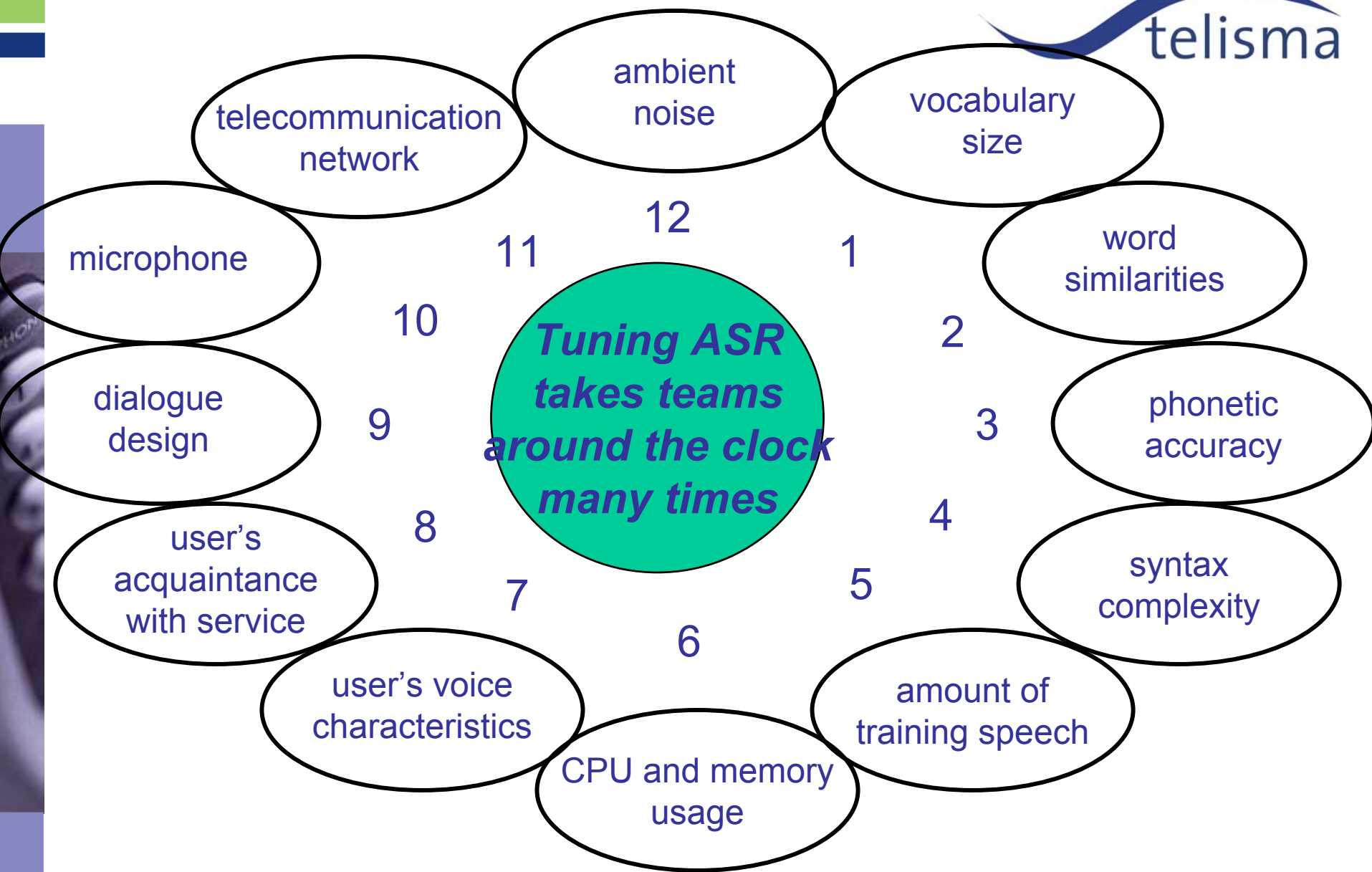
Overview of ASR components



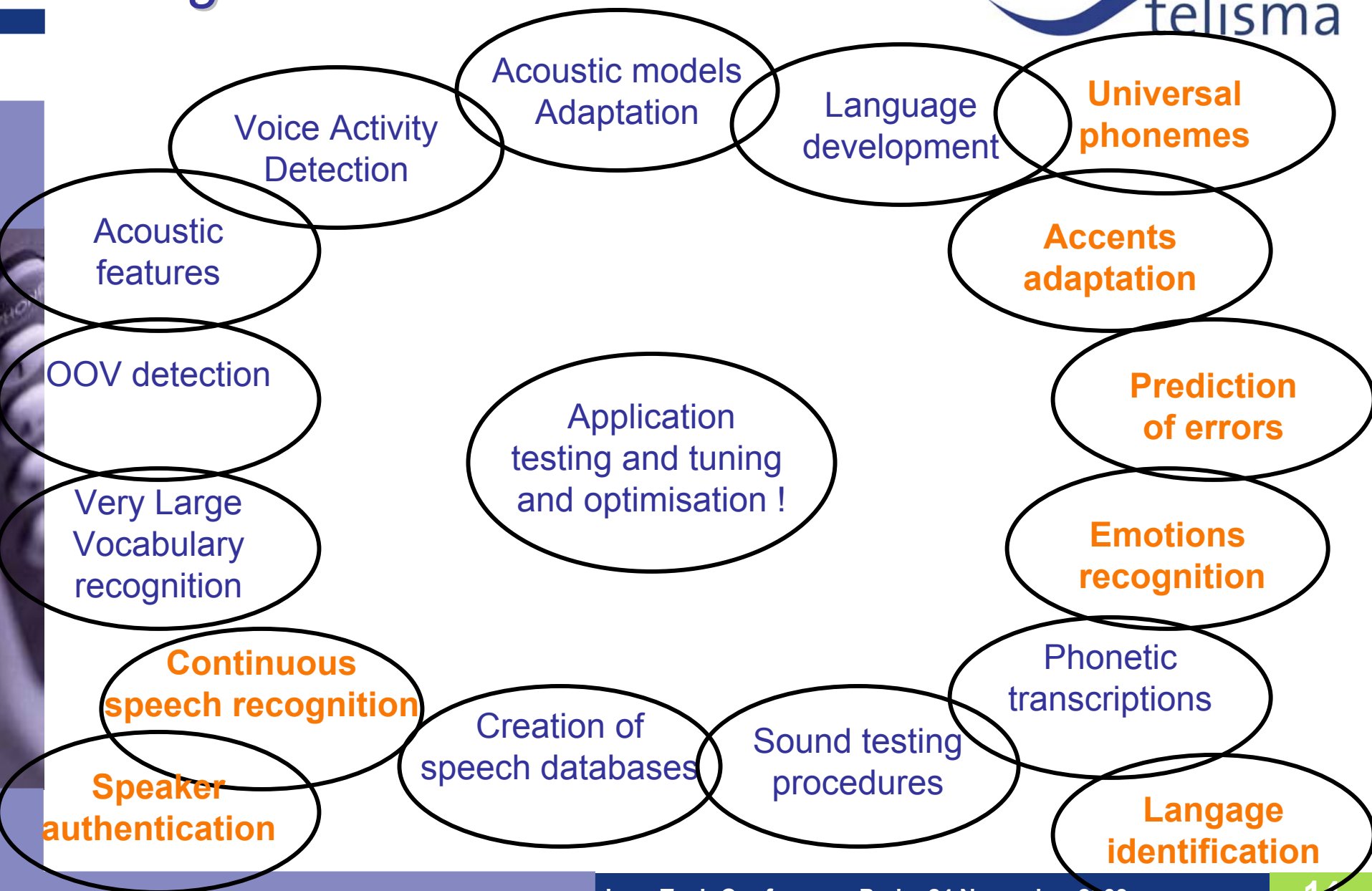
Factors impacting ASR performance



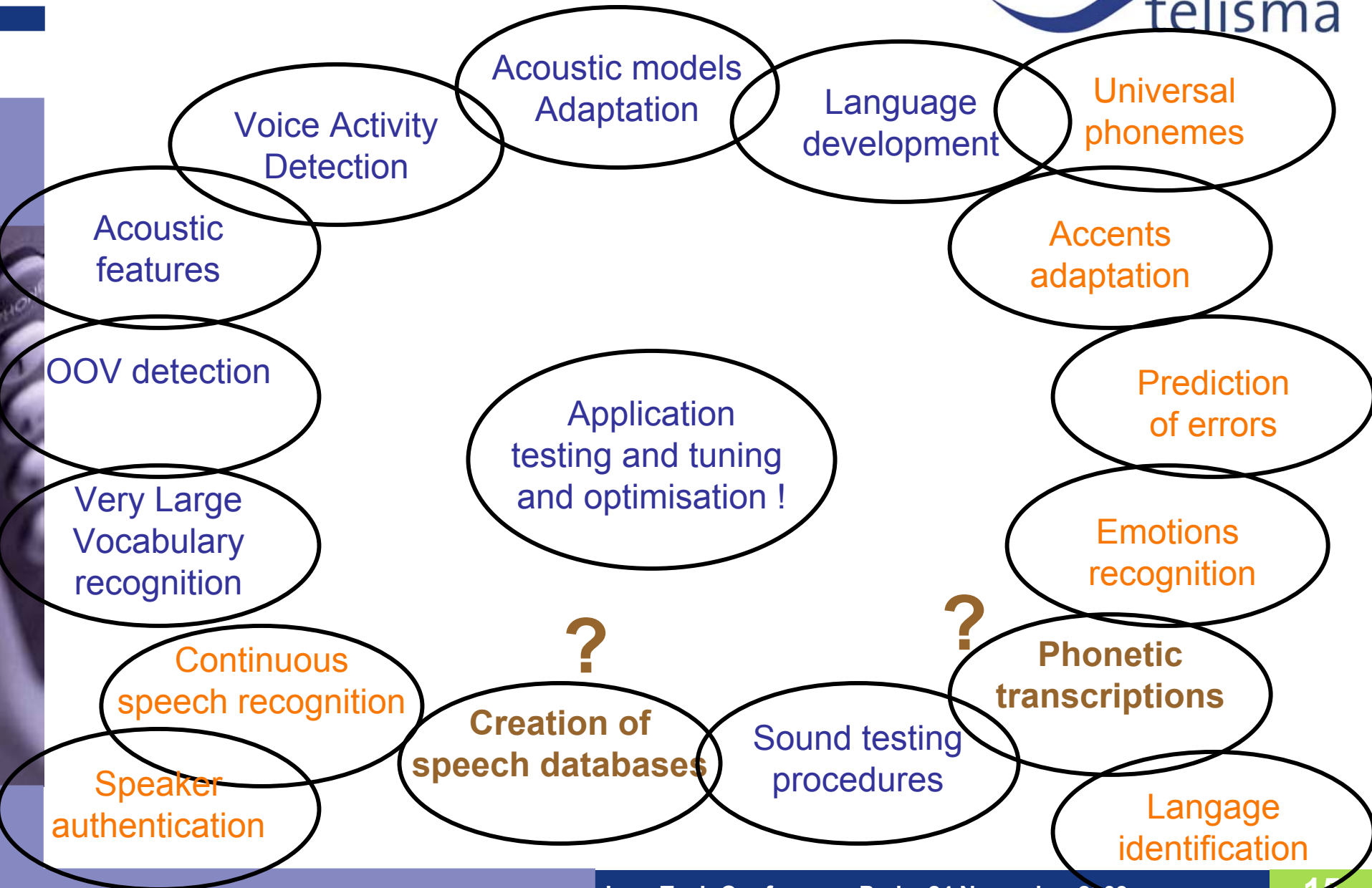
Factors impacting ASR performance



Things nice and useful to have



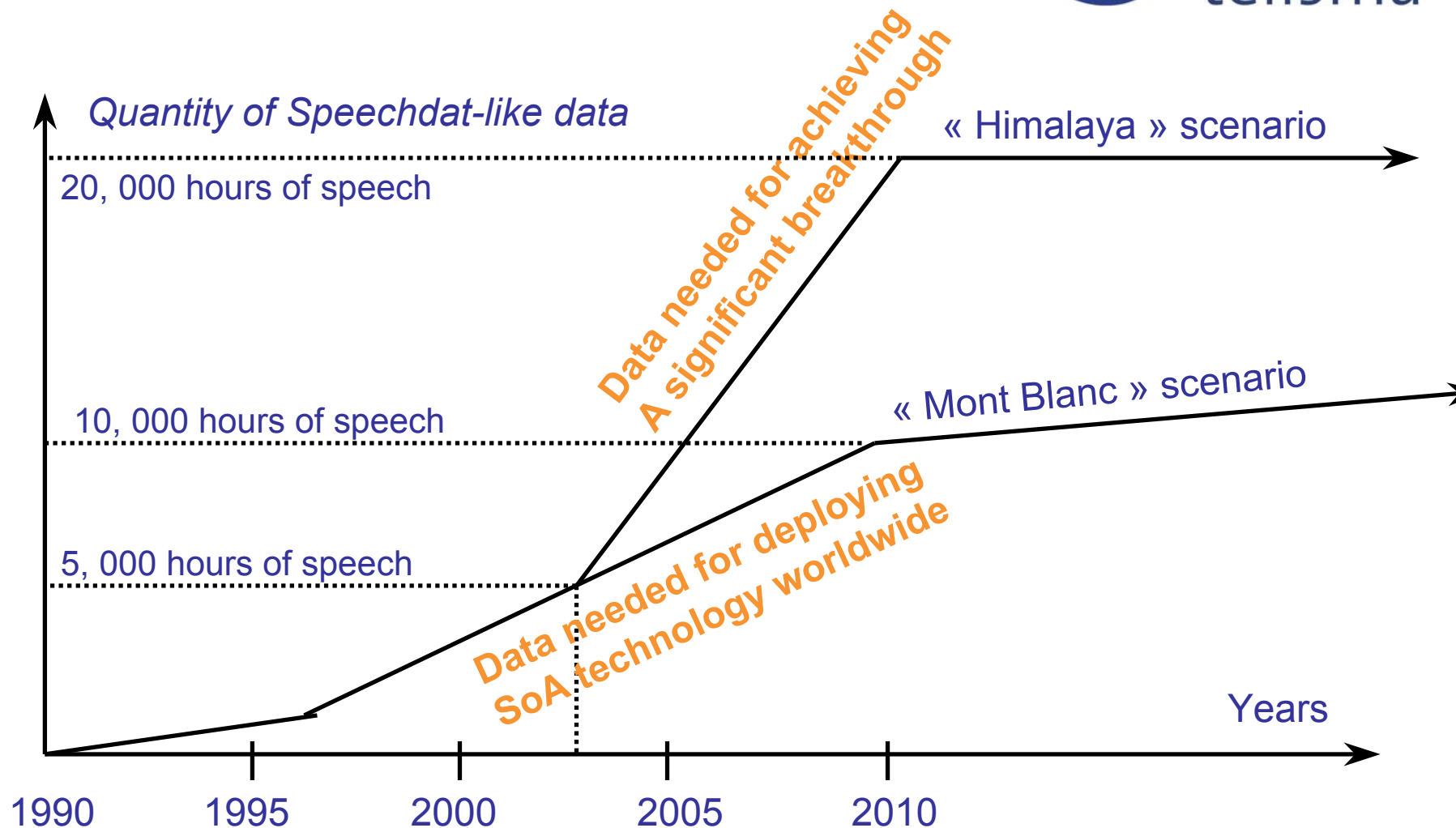
What about these “must have”s ?



- Until now, we believed that accurate phonetic transcriptions, taking into account multiple pronunciations, are mandatory to achieve best ASR accuracy
- But this is questioned by recent findings :
 - Manually verified transcription do not always overperform automatic ones
 - Grapheme acoustic models perform almost as well as phoneme models

- Never forget : field data (specific data) is good for ASR
- What about generic data ? (Switchboard, Callhome, Speechdat, ...)
- Until now, I believed that we (the speech industry) had to more aggressively climb up the speech data wall, moving from today's "Mont Blanc" scenario to a "Himalaya" scenario :
 - **"Mont Blanc" scenario**
Continue recording 1000 speakers databases (approximately 20 hours of speech) for another 50 languages and for fixed and mobile networks
 - **"Himalaya" scenario**
Speed up language databases expansion by recording much more data, up to 10 000 speakers (200 hours of speech) for main languages in the world
- But this is questioned by recent extrapolations: increasing speech data 100 times may not even halve error rates
 - Cf. EUROSPEECH 03 paper by Roger Moore

Climbing up the speech data wall

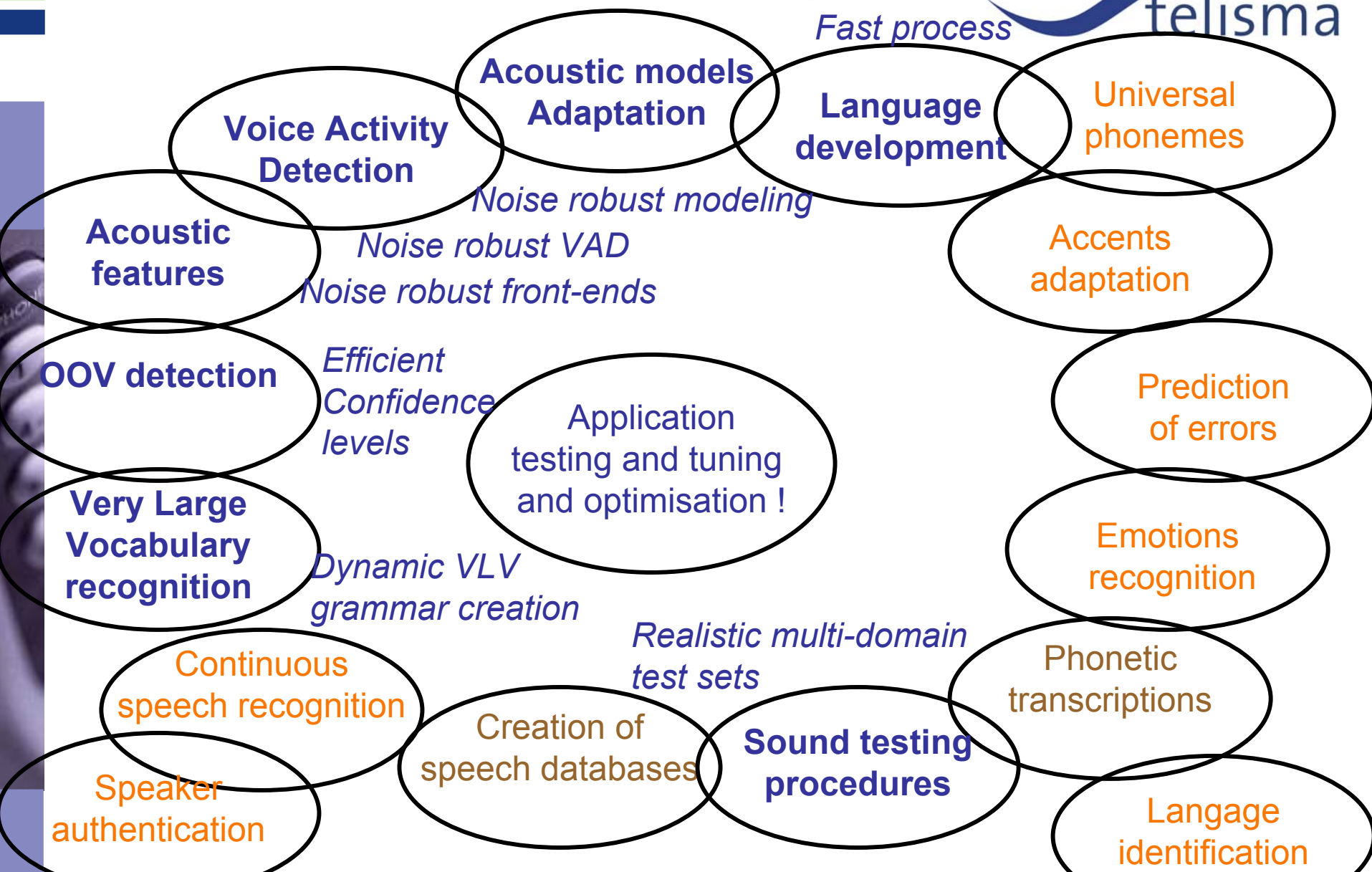


Speech databases : more design, less brute force

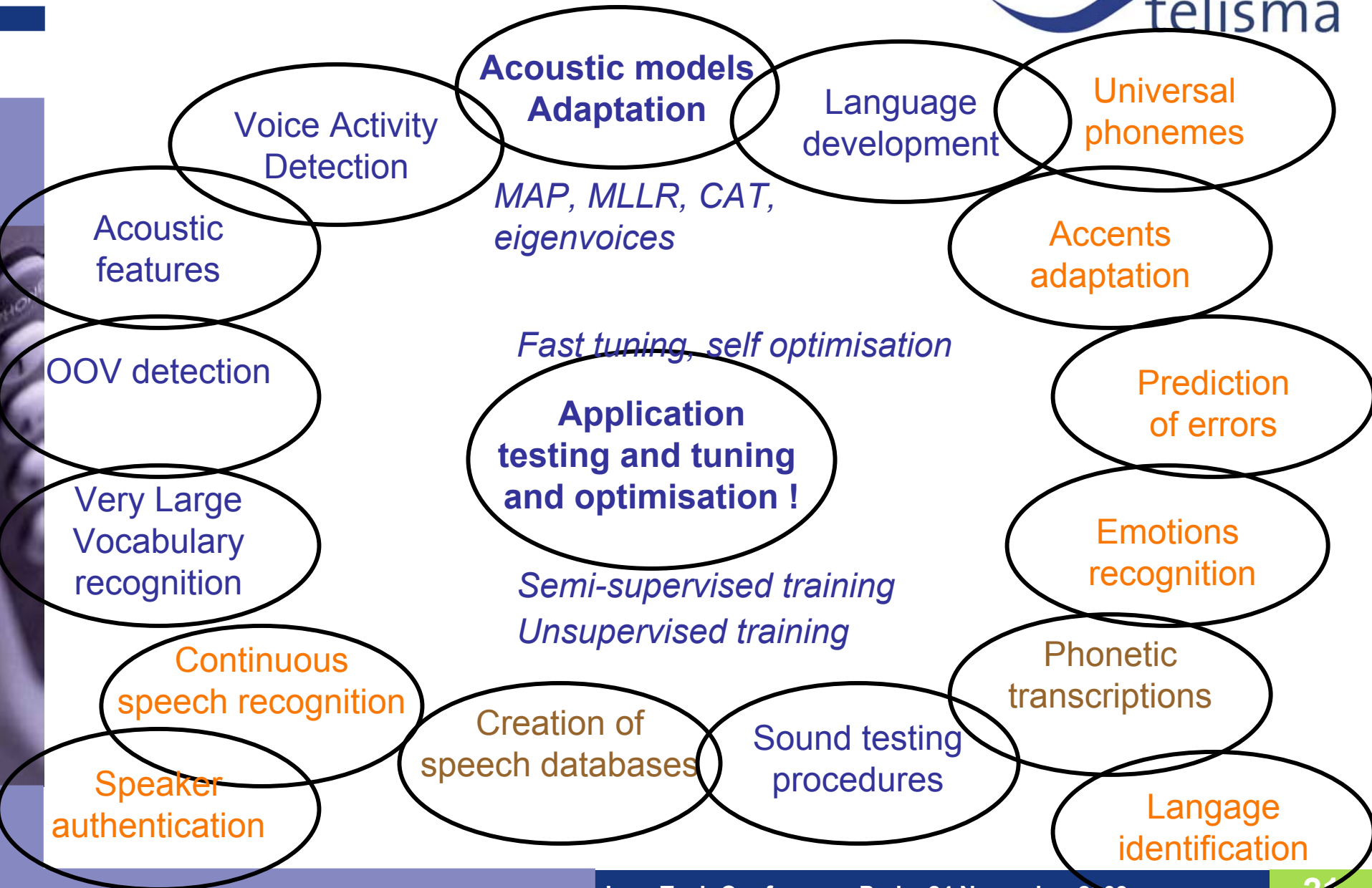


- Shared cost speech databases projects are beneficial
- But brute force (pure accumulation) might not be necessary
- In the wake of the SpeechDat success stories, let us now design new more structured databases meeting the needs of new algorithmic approaches
- For example, we have recently launched a French Speech Database project, codename NEOLOGOS, aimed at the creation of structured databases adapted to clustering adaptation and eigenvoices techniques.
- Can we think of a multilingual databases enabling real mutualisation of speech data across languages and across main relevant factors such as network and noise conditions ?

The ASR "must have"s components



The ASR "keystones" components



More cooperation is needed between the ASR players

*Focus on technology
priorities*

*Cooperate more with
the industry*



Feedback to researchers

**ASR research
community**

**ASR
industry**

ASR customers

Set customers' expectation

*Build partnerships with
the ASR industry*

*Accept to contribute
more to the ASR
industry effort*

Concluding remarks



- Like it or not, efficient ASR remains expensive
- Active collaboration with customers is a key factor
- Much investment remains to be made, even on KEYSTONES and MUST HAVE components
- The speech scientific community is very creative and active !
- Closer collaboration between the speech scientific community and with the industrial partners can only be beneficial
- ASR must still develop a significant effort in pursuing two A goals:
 - Become more Accurate
 - Become more Automatic