

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu

## Synthetic Speech Intelligibility

João Miguel Ferreira <sup>1, 2</sup>

Marina Vigário <sup>2</sup>

Fábio Fernandes <sup>1</sup>

Fernanda Belchior <sup>1</sup>

Susana Azevedo <sup>1</sup>

Ana Rita Neca <sup>1</sup>

<sup>1</sup> Escola Superior de Saúde Instituto Politécnico de Setúbal

<sup>2</sup>Laboratório de Fonética do Centro de Linguística da Universidade de Lisboa



Lisboa 2014

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

## Porquê estudar Inteligibilidade em Fala Sintetizada?

- Diversas patologias impedem o uso funcional da fala/voz, pelo que a fala/voz sintetizada é a alternativa mais funcional. A qualidade da mesma é fundamental para a eficácia comunicativa, sendo que a inteligibilidade é uma das medidas de qualidade.
- A qualidade da fala/voz sintetizada é ainda fundamental para o desenvolvimento linguístico dos utilizadores, tendo em consideração ser o modelo de aprendizagem dos mesmos.

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

## Inteligibilidade em Fala Sintetizada

- Braga, Freitas, & Barros, (2002) referem-se à inteligibilidade da fala sintetizada como sendo um grau de compreensão de fala, sendo que a inteligibilidade demonstra **a capacidade que o ouvinte tem de identificar palavras** ou frases sintetizadas. Esta identificação refere-se a palavras e frases apresentadas isoladamente sem contexto linguístico (Ralston, Pisoni & Mullennix, 1989, citados por Koul & Clapsaddle, 2006).



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

## Inteligibilidade em Fala Sintetizada

- O estudo de Logan, Greene & Pisoni ( 1989 cit. por Winters & Pisoni, 2004) revelou resultados bastante semelhantes entre a fala natural e sintetizadores de fala de alta qualidade (DECTalk) quando se usava a resposta fechada, e significativamente piores quando utilizavam a versão de resposta aberta.
- Quando se utilizavam sintetizadores de baixa qualidade a diferença era bastante significativa em ambos os estudos
- Um estudo para o PE utilizado a versão portuguesa do Dynamic Rhyme Test (DRT), comparando resultados entre jovens adulto e idosos, com voz sintetizada (voz Célia Acapela), apontava para taxas de acerto na ordem dos 90% (Ferreira; Laranjeira; Perpétua; 2012).

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

## Inteligibilidade em Fala Sintetizada

- O DRT é aceite como um dos *standards* para avaliação da inteligibilidade, utilizando pares de palavras.
- O DRT ou o Modified Rhyme Test (MRT) apesar de apresentarem diversas vantagens como a fiabilidade, facilidade de aplicação em participantes inexperientes e facilidade de cotação (Logan, Pisoni Greene 1985 cit por Koul, 2003), bem como as suas versões portuguesas, não têm em consideração a estrutura silábica do PE, nem o acento, nem o tamanho das palavras.

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Experiência Inteligibilidade

Tendo em consideração as fragilidades dos instrumentos utilizados, dada a sua não adequação ao PE, elaborou-se um novo instrumento, pensado para o PE, que será descrito mais à frente.

Dada a falta de investigação relativa à qualidade dos SSF que suportam estas vozes, optamos por iniciar uma investigação a diversos aspetos das mesmas. Numa primeira fase a avaliação centra-se na Inteligibilidade, sendo que posteriormente nos focaremos em aspetos suprasegmentais (prosódia).

• Introdução

• Estudo



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Experiência Inteligibilidade

Comparar 3 vozes sintetizadas\* entre si e com uma voz natural, relativamente a taxas de acerto e tempos de reação

### Hipótese

Dado a voz eletrónica utilizada ser uma voz de alta qualidade, não é esperada diferença significativa na inteligibilidade.

\*As 3 vozes foram escolhidas por serem as que são mais utilizadas nos SSF do PE, em software de CAA respetivamente Voz Célia (Acapela Group, GRID II), Voz Joana (Nuance, Nuance Vocalizer) e Voz Madalena (Nuance, Software Comunicar com Símbolos) (Ferreira 2013)

• Introdução

• Estudo



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Objetivos

Contribuir para a melhoria da inteligibilidade dos sistemas de voz sintetizada.

Comparando as vozes dos sistemas de síntese de fala para o Português Europeu, entre si e com uma voz natural

Averiguar se o tipo de voz tem influência na taxa de acertos e tempos de reação

• Introdução

• Estudo





# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Experiência Inteligibilidade Instrumento de Avaliação

- Teve-se por base o modelo dos testes *standard* de avaliação da inteligibilidade de síntese de fala.
- Elaborou-se um novo instrumento tendo em conta o tamanho da palavras (nº de sílabas), estrutura silábica e e distribuição de acento do PE. Não é tida em consideração a frequência dos segmentos no PE.
- Consulta da base de dados FrePOP 1.0 (Frota; Vigário; Martins & Cruz 2010) para as frequências de ocorrência no PE.
- Utilização da ferramenta FreP 3.1.1.1 (Martins ; Vigário& Frota ;2011) para verificação das frequências no instrumento.

• Introdução

• Estudo

• Metodologia



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Instrumento de Avaliação

### Localização do Acento

	Frequência no PE	Instrumento
<b>Final</b>	18%	19%
<b>Penúltima</b>	58%	56%
<b>Antepenúltima</b>	2%	2%
<b>Monossilábicas</b>	22%	23%

### Número de Sílabas

	Frequência no PE	Instrumento
<b>Monossilabos</b>	22,9%	22,9%
<b>Dissílabos</b>	41,0%	40,8%
<b>Trissílabos</b>	22,9%	23,6%
<b>Polissílabos</b>	13,3%	12,7%

### Tipos Silábicos

157 pseudopalavras  
157 palavras prosódicas  
358 sílabas

	Frequência no PE	Instrumento
<b>CV</b>	42%	44%
<b>V</b>	16%	13%
<b>CVC</b>	11%	11%
<b>CVGN</b>	6%	4%
<b>CVN</b>	5%	5%
<b>VC</b>	3%	4%
<b>CCV</b>	2%	3%
<b>% Total de Tipos</b>	85%	85%

• Introdução

• Estudo

• Metodologia

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

• Estudo

• **Metodologia**

## Metodologia

- *Nº de sujeitos=80 18-25 anos*
- *PE língua materna*
- *S/ alterações auditivas (despiste auditivo prévio)*
- *4 grupos de 20 sujeitos distribuídos pelas 4 vozes*
- *4 vozes*
  - *3 de síntese de fala Célia , Madalena e Joana*
  - *1 voz natural (feminina)*
- *Cada grupo sujeito a uma só condição experimental*
- *Apresentação da experiência em SuperLab 4.5*

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

• Estudo

• **Metodologia**

## Metodologia

- Foi pedido aos participantes que ouvissem os estímulos (pseudopalavras) e que imediatamente a seguir as escrevessem o mais adequadamente possível, na caixa de texto reservada para o efeito.
- Os estímulos eram apresentados pelo SuperLab, através de *headphones* e os participantes digitaram as respostas no computador.

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Variáveis

- Participantes
- Sistemas de Síntese de Fala (SSF) /Vozes
- Taxa de Acerto Global (totais entre vozes)
- Tempo de Reação por tipo de SSF
- Taxas de Acerto Nº Sílabas por tipo de SSF
- Taxas de Acerto Acento por tipo de SSF

• Introdução

• Estudo

• Metodologia



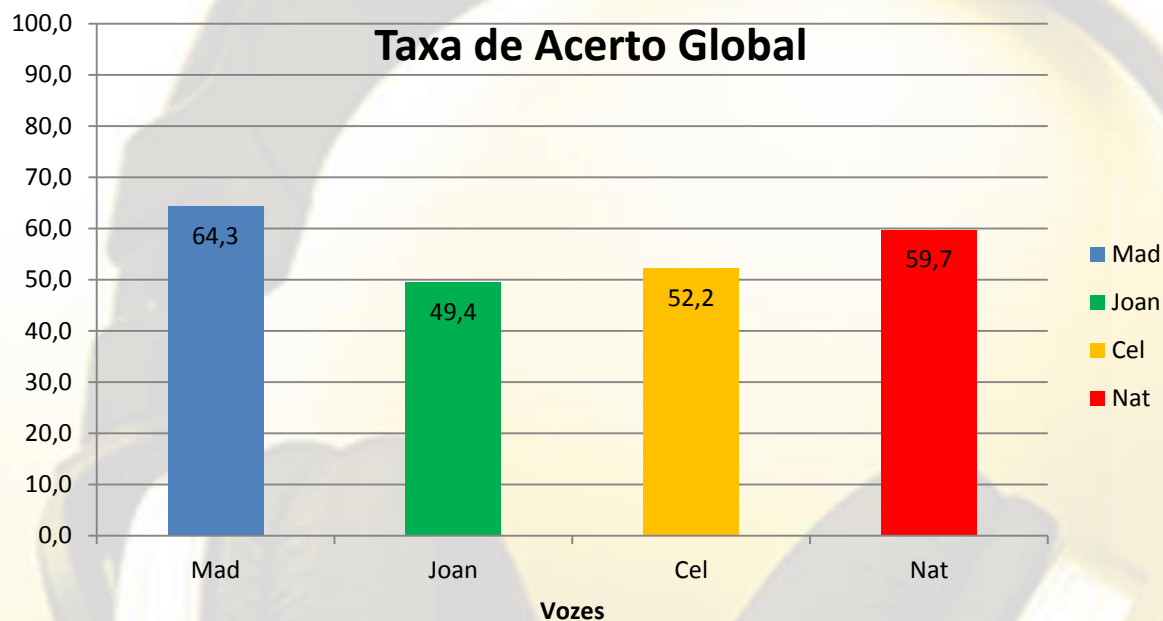

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



- Introdução
- Estudo
- Metodologia
- Resultados

/ LabFlon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



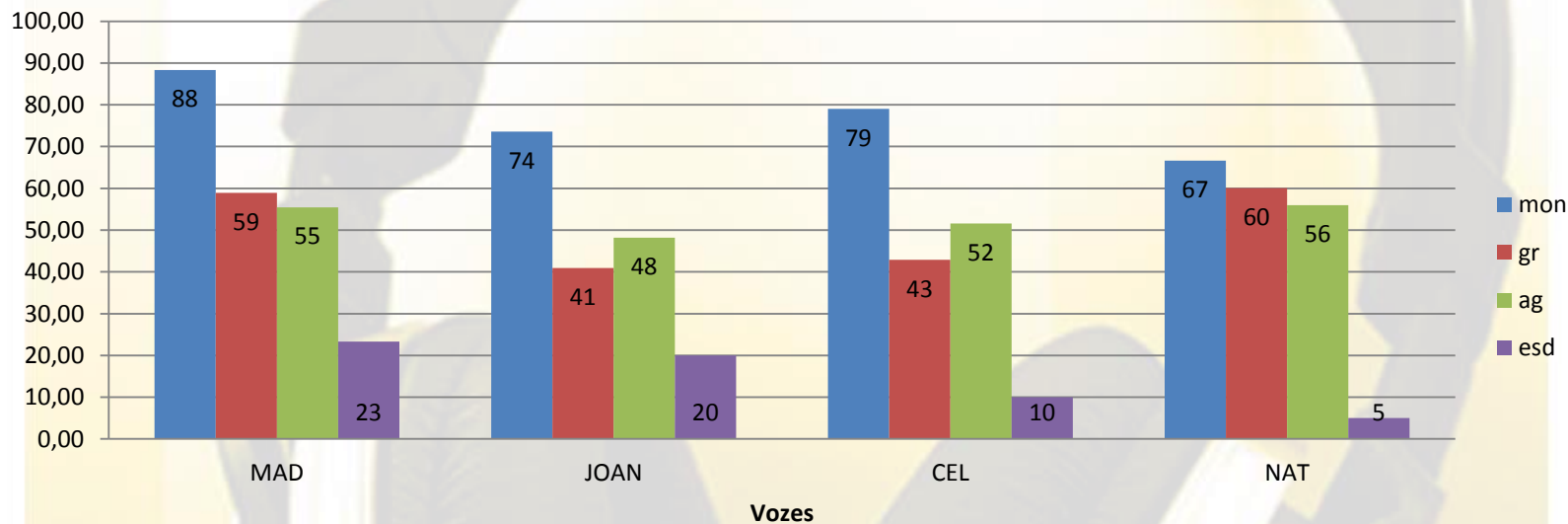
• Introdução

• Estudo

• Metodologia

• Resultados

## Taxa de Acerto por padrão acentual




# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu

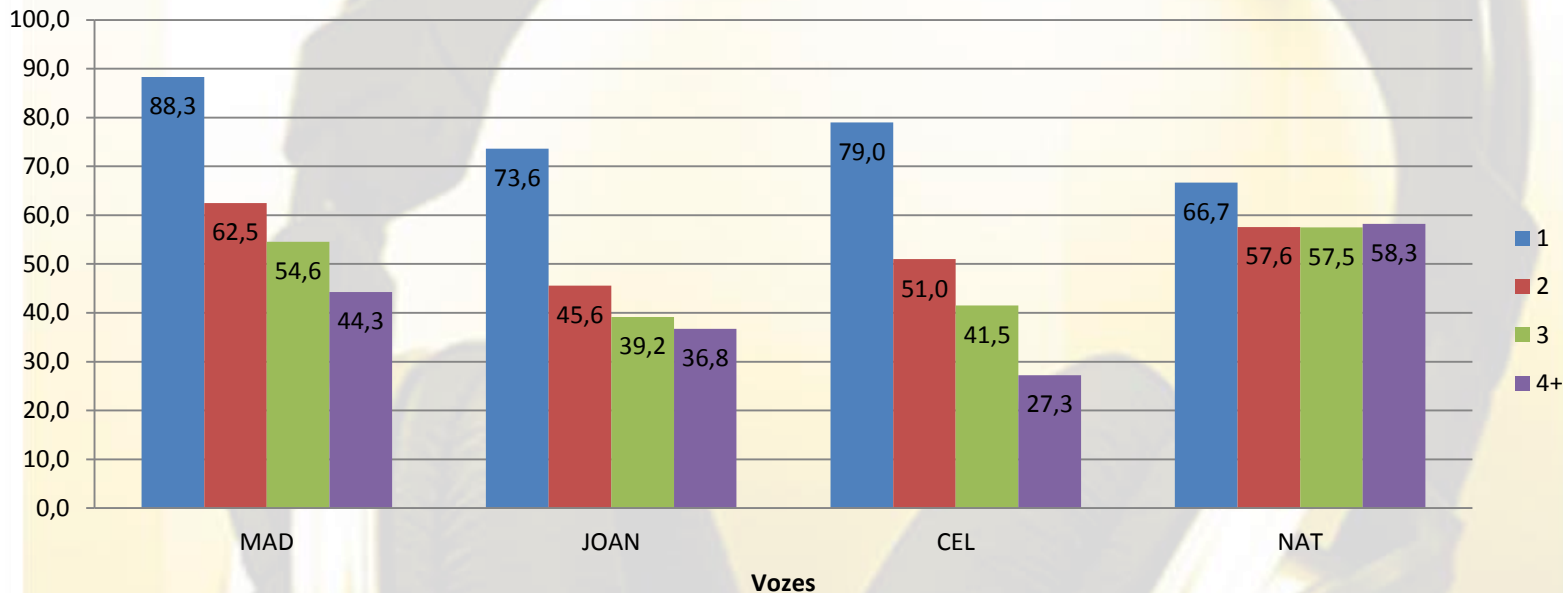


- Introdução
- Estudo
- Metodologia
- Resultados

/ LabFlon /



## Acerto por Nº de Sílabas





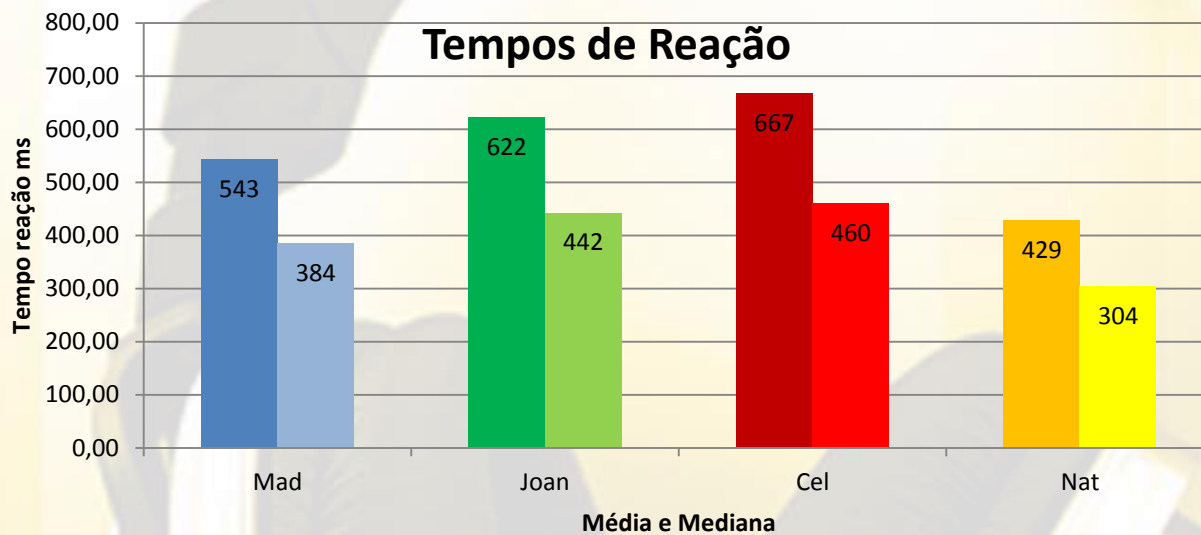

# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



- Introdução
- Estudo
- Metodologia
- Resultados

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Discussão

- Podemos verificar que a voz Madalena apresenta resultados globais superiores às outras vozes e inclusive à voz Natural. Estes resultados, contrariam a maioria dos estudos realizados para outras línguas.
- A voz Natural apresenta resultados muito consistentes e sempre acima dos 55%, independentemente do tamanho dos estímulos apresentados.
- Ainda relativamente ao tamanho dos estímulos, nas pseudopalavras monossilábicas, todas as vozes sintetizadas conseguem taxas de acerto superiores às da voz natural, o que poderá indicar que segmentalmente as vozes sintetizadas apresentam uma qualidade muito boa.

• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Discussão

- A voz Natural apresenta um tempo de reação mais baixo do que qualquer uma das vozes sintetizadas, sendo que este resultado era esperado e verificado em estudos para outras línguas (Rajinder Koul & Dembowski, 2010)

• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Conclusões

- As vozes sintetizadas apresentaram resultados bastante próximos da voz natural ( e em alguns casos superiores), pelo que se pode considerar que estas vozes apresentam uma qualidade bastante satisfatória.
- Os dados preliminares da análise qualitativa do tipo de erro parecem indicar que o acento é o principal fator de erro, com por exemplo as palavras esdrúxulas com uma taxa de erro muito alta.
- Tendo isto em consideração, o acento será alvo de avaliação mais detalhada em estudos futuros.
- Poderá haver diferença entre perceção/identificação do local do acento e a correta marcação ortográfica do acento

• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão

• Conclusões

/ Lab|Flon /



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão

• Conclusões

- Em todos os SSF/vozes a taxa de acerto vai reduzindo conforme aumenta o número de sílabas, o que não sucede com a voz natural. Nos polissílabos isto é notório, sendo que o facto de estes representarem somente cerca de 13% das ocorrências no PE (FrePOP), poderá ter influência nos resultados.
- Os resultados deste estudos, parecem indicar que a tarefa é mais complexa do que a tarefa utilizada (resposta aberta vs resposta fechada) em estudos anteriores para o PE e mesmo em estudos semelhantes para o Inglês (Ferreira, Laranjeira & Perpétua; 2012)



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão

• Conclusões

- O instrumento da inteligibilidade de avaliação ainda carece de um estudo mais aprofundado, nomeadamente quanto a valores “normativos”, pelo que a utilização de mais vozes naturais será um fator a considerar



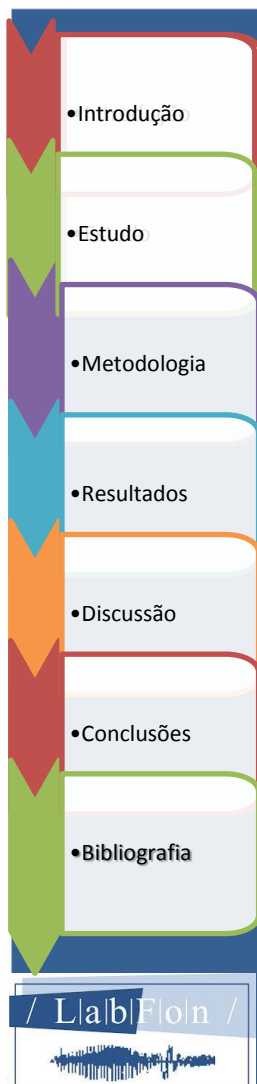
# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Bibliografia

- Braga, D., Freitas, D., & Barros, M. J. (2003). A DRT approach for subjective evaluation of intelligibility in European Portuguese synthetic speech. *WSEAS Transactions on Computers*, 2(2), 510–514.
- Ferreira, J., Laranjeira, M., & Perpétua, P. (2012). Inteligibilidade da Fala Sintetizada no Português Europeu. *Unpublished*, 39.
- Ferreira, J. (2013) Inquérito online à utilização de Sistemas de Síntese de Fala para o Português Europeu em pessoas com Necessidades Complexas de Comunicação, *unpublished*
- Frota, S., Vigário, M., Martins, F., & Cruz, M. (2010). FrePOP Database- Frequency of Phonological Objects of Portuguese. Laboratório de Fonética- FLUL- CLUL.
- Koul, R. (2003). Synthetic Speech Perception in Individuals With and Without Disabilities. *Augmentative and Alternative Communication*, 19(1), 49–58. doi:10.1080/0743461031000073092
- Koul, R., & Clapsaddle, K. C. (2006). Effects of repeated listening experiences on the perception of synthetic speech by individuals with mild-to-moderate intellectual disabilities. *Augmentative and alternative communication (Baltimore, Md. : 1985)*, 22(2), 112–22. doi:10.1080/07434610500389116
- Martins, F., Vigário, M., & Frota, S. (2011). FreP- Frequency in Portuguese.
- Winters, S. J., & Pisoni, D. B. (2004). *Perception and Comprehension of Synthetic Speech - Progress Report nº26* (Vol. 26, pp. 95–138).



# Inteligibilidade em Voz Sintetizada

Avaliação de Sistemas de Síntese de Fala em Português Europeu



## Agradecimentos

À Susana Correia pela ajuda inicial na montagem da experiência no SuperLab

Aos participantes no estudo

• Introdução

• Estudo

• Metodologia

• Resultados

• Discussão

• Conclusões

• Bibliografia

/ LabFlon /

