Teste 1

1. Considere a palavra “recado”. Se na sequência fonética trocarmos apenas a característica de vozeamento de cada uma das 2ª e 3ª consoantes qual seria a sequência textual que se obteria? Justifique.
2. A prosódia é constituída por 3 componentes. Indique-as e explique os seus papéis na criação de um padrão prosódico relativo ao foco de uma frase.
3. Explique o funcionamento glotal na produção dos sons vozeados não-oclusivos.
4. Um dos blocos mais importantes do conversor texto-fala é o conversor fonético. É um módulo que se apoia numa representação codificada dos fonemas. Indique e distinga dois códigos utilizados para a representação fonética e faça a distinção conceptual entre fone e fonema.
5. Explique porque razão uma pessoa surda de nascimento apresenta maior dificuldade que o normal em adquirir a competência da fala.
6. A síntese da fala pode realizar-se entre outros processos por meio de concatenação de segmentos temporais de fala original minimamente processados. Mostre em que consiste um difone e compare com uma sílaba, explicando os problemas existentes em cada um dos casos.
7. A correcção ortográfica é uma fase do processamento da fala destinado à síntese que se realiza na fase inicial do processo. Identifique essa fase inicial e descreva duas outras operações tipicamente realizadas nessa fase inicial.
8. É sabido que as ressonâncias do tracto vocal formam o timbre de muitos dos sons da fala. Mostre, como se calcularam os valores das frequências dos formantes da vogal 6 (SAMPA).
9. Explique em que consiste o triângulo das vogais e as variáveis que o estruturam e esboce-o.
10. Um espectrograma de banda larga deverá mostrar diferentes gamas de frequências do espectro da fala que relativamente a um espectrograma de banda estreita? Justifique a resposta. Qual é o mais adequado à observação da frequência fundamental da fala? Justifique.

Teste 2

1. Considere a palavra “ficas”. Se na sequência fonética invertermos apenas a característica de vozeamento de cada uma das 2 primeiras consoantes qual seria a sequência textual que se obteria?
2. A produção oral das vogais é associada a ressonâncias do tracto vocal. As respectivas frequências denominam-se formantes. Explique resumidamente como podem as frequências formantes ser utilizadas para distinguir as vogais e interprete a representação gráfica do triângulo das vogais.
3. Explique em que consistem os espectrogramas de banda larga e de banda estreita e para que aspectos da observação do sinal de fala, por exemplo de tipo vozeado, são mais adequados.
4. Explique em que consiste a prosódia da fala humana e quais são as variáveis específicas que a caracterizam.
5. Na síntese da fala é necessário preparar o texto para ser convertido. A primeira fase dessa preparação denomina-se pré-processamento. Indique dois dos processos que são realizados nesta fase sobre o referido texto.
6. A laringe e nomeadamente a glote, têm uma função essencial na produção da fala. As cordas vocais permitem a produção de sons com excitação periódica denominados vozeados ou sonoros. Explique o funcionamento glotal na produção destes sons.
7. Um dos blocos mais importantes do conversor texto-fala é o conversor fonético. Os códigos SAMPA e IPA são utilizados para a representação fonética. Descreva duas características diferenciadoras destes códigos.
8. A audição tem um relacionamento muito importante com a fala. Explique qual é este relacionamento.
9. Explique em que consiste e a que finalidade se destina a anotação de um sinal de fala.

Teste 3

1. Distinga um difone de uma sílaba no âmbito da síntese da fala.
2. Considere a palavra “feridas”. Se na sequência fonética trocarmos apenas a característica de vozeamento de cada uma das 1ª e 3ª consoantes qual seria a sequência textual que se obteria? Justifique.
3. Explique em que consiste a frequência fundamental da fala, f0 e qual a forma da sua utilização para assinalar uma interrogação.
4. Explique o ciclo glotal na produção dos sons vozeados não-oclusivos.
5. Um dos blocos mais importantes do conversor texto-fala é o conversor fonético. Apresente e explique 3 situações em que uma mesma letra tem transcrições fonéticas diversas.
6. Um espectrograma de banda estreita deverá mostrar maior detalhe dos formantes ou de f0? Justifique a resposta.
7. Uma das características importantes da audição humana é a existência de bandas críticas. Explique em que consiste uma banda crítica.
8. Indique e explique as principais tarefas a realizar na fase do pré-processamento do texto destinado à síntese da fala.
9. Explique em que consiste um formante da fala e como se pode apreciar a sua existência através do espectrograma da fala.
10. Explique como se pode organizar o domínio dos sons das vogais em função das frequências formantes.

Teste 4

1. Considere a palavra “javali”. Se na sequência fonética trocarmos apenas a característica de vozeamento das 2 primeiras consoantes qual seria a sequência textual que se obteria:

a) xabali.
b) zavali.
c) fazali.
d) xafali.

2. A produção oral da vogal [α] como na palavra “nas” é associada a ressonâncias do tracto vocal. No caso do comprimento efectivo deste tracto ter o valor de 17 cm a mais baixa frequência de ressonância terá o valor aproximado de:

a) 504 Hz
b) 2016
c) 1008
d) Todas as alíneas anteriores estão erradas.

3. A correcção da resposta em frequência no gerador de sinal sintético de voz baseado em modelos paramétricos, denominada correcção de radiação dos lábios deve-se:

a) à forma arredondada da abertura dos lábios
b) à radiação acústica da boca, mas não é necessária na maior parte dos casos
c) ao tamanho da abertura bucal e consiste num passa-alto
d) todas as alíneas estão certas

4. Na zona de valores elevados de taxa de passagens por zero e valores médios da energia média encontram-se sinais relativos a:

a) silêncio com “offset”.
b) vozeados.
c) silêncio.
d) Não vozeados.

5. Na determinação de f0 de um segmento de sinal de voz utilizou-se a técnica da autocorrelação e detecção de picos. Para tal deveria ter-se utilizado:

a) Uma janela de duração aproximadamente igual ao valor médio do período a medir e passo igual
b) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual
c) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual ao período a medir
d) Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo menor que o menor período a medir

6. O espectrograma de um sinal de voz permite uma observação muito rica das características do mesmo, nomeadamente no respeitante a amplitudes, conteúdos frequenciais e variação temporal. No processo de determinação do espectrograma que decorreu na aula prática, comprovou-se que:

a) A duração da janela temporal utilizada influenciava fortemente a resolução frequencial obtida sendo melhor a resolução obtida com janela mais extensa.
b) O passo a utilizar deveria ser tanto mais pequeno quanto melhor a resolução frequencial desejada.
c) A duração da janela temporal utilizada não influenciava sensivelmente a resolução frequencial obtida.

d) A duração da janela temporal utilizada influenciava fortemente a resolução frequencial obtida sendo pior a resolução obtida com janela mais extensa.

7. Na síntese de sinal de voz por concatenação deve ser evitado o seguinte:

a) Efectuar a manipulação prosódica do sinal resultante.

b) Utilizar segmentos que tenham características prosódicas muito semelhantes ou próximas dos valores pretendidos devido a possibilidade de reverberação.

c) Utilizar valores iniciais nos filtros LPC nas mudanças de quadro (frame), quando for o caso

d) Associar segmentos temporais recortados de zonas diferentes do material de voz de base.

8. Sabendo que as frequências formantes do sinal de fala indicam certas características do tracto vocal que o produziu, podemos determinar essas frequências:

a) Por meio dos coeficientes LPC do sinal de excitação glotal.

b) De forma directa do cepstro do sinal de voz.

c) Através da determinação dos picos do espectro do sinal de voz determinados com janela de comprimento adequado.

d) Todas as outras alíneas estão erradas.

9. Na determinação dos N coeficiente LPC pelo método da auto-correlação, além do termo unitário, de um sinal de voz é necessário resolver um sistema de N equações denominadas normais e:

a) Os termos independentes deverão ser calculados separadamente.

b) Para construir esse sistema de equações é necessário calcular N+2 termos de correlação

c) Para construir esse sistema de equações é necessário calcular N+1 termos de correlação

d) Para construir esse sistema de equações é necessário calcular N termos de correlação

10. Os principais métodos para determinar a frequência fundamental de um sinal de fala são:

a) O cepstro, a autocorrelação, a AMDF, as energias ou amplitudes médias e pelo menos 10 coeficientes LPC.

b) Os da alínea a) acrescidos do espectrograma

c) Os da alínea a) exceptuando o cepstro

d) Todas as outras alíneas estão erradas

11. Descreva um sistema de conversão texto-fala, que recebe texto ASCII e produz um sinal digital em formato .WAV, baseando-se num diagrama de blocos adequado e explicando com exemplificação as operações necessárias nesse diagrama, mencione também as tarefas e recursos de base eventualmente necessários para a operação efectiva do sistema além dos algoritmos que estarão compreendidos no diagrama.

12. Explique o conceito de “pulso glotal” e compare-o com o sinal de erro da modelização LPC de sinais de fala.

Teste 5

1. Em relação às dificuldades técnicas presentes nas diferentes aplicações do reconhecimento automático de fala, é correcto afirmar:

- □ O reconhecimento de palavras isoladas apresenta pequena dificuldade mesmo em ambientes acústicos muito ruidosos (SNR muito baixo).
- □ O reconhecimento de fala contínua quando a aplicação é dependente do falante é geralmente mais difícil do que quando é independente do falante.
- □ O reconhecimento de fala contínua torna-se mais fácil diminuindo a perplexidade linguística da tarefa de reconhecimento.
- □ Todas as respostas anteriores são incorrectas.

2. Nos sistemas de reconhecimento de fala , o módulo de Análise:

- □ deveria, idealmente, extrair apenas a informação discriminante para a tarefa de reconhecimento.
- □ transforma, em geral, o sinal acústico de fala numa sequência de vectores de características.
- □ deve conduzir a uma representação onde a variabilidade em cada classe é relativamente pequena e a separação entre classes é relativamente grande.
- □ Todas as respostas são correctas.

3. Nos sistemas de reconhecimento de fala, o módulo de Classificação:

- □ baseia-se frequentemente numa estrutura que integra numa única "rede" a informação relativa ao modelo acústico e ao modelo linguístico.
- □ integra o módulo de Análise do sinal quando a tarefa de reconhecimento é dependente do falante.
- □ integra o módulo de Análise do sinal quando se utilizam modelos acústicos, correspondentes às unidades linguísticas elementares, definidos ao nível da palavra.
- □ Todas as respostas anteriores são incorrectas.

4. No sentido de aumentar a capacidade de generalização dos sistemas de reconhecimento automático de fala, em geral:

- □ aumenta-se o mais possível o número de iterações de treino e o número de parâmetros livres, independentemente do tamanho da base de dados para treino.
- □ diminui-se o mais possível o número de parâmetros livres do sistema, independentemente do tamanho da base de dados para treino e do número de iterações de treino.
- □ durante o processo de treino utiliza-se um conjunto de dados, ainda não observados pelo sistema, para acompanhar a evolução da capacidade de generalização e agir em conformidade.
- □ termina-se o processo de treino quando o erro estimado sobre o conjunto de treino é mínimo, desde que o número de parâmetros livres do sistema seja suficientemente elevado.

5. Seja **X** uma sucessão de vectores de características extraídos de um segmento de sinal de fala que corresponde, por hipótese, à classe **W**. O módulo de Classificação de um sistema de reconhecimento de fala calcula o valor:

- □ da probabilidade *a priori*, $P(\mathbf{X}|\mathbf{W})$, através de um modelo Linguístico.
- □ da verosimilhança acústica, $P(\mathbf{X}|\mathbf{W})$, através de um modelo Acústico.
- □ da probabilidade *a posteriori*, $P(\mathbf{W})$, através de um modelo Linguístico.
- □ Todas as respostas anteriores são incorrectas.

6. Seja **X** uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja $\mathbf{W}_j$, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um reconhecedor de fala baseado no critério Bayesiano de classificação reconhece **X** como pertencente à classe $\mathbf{W_C}$ se e só se:

- □ $P(\mathbf{X}, \mathbf{W_C})\, P(\mathbf{W_C}) > P(\mathbf{X}, \mathbf{W_j})\, P(\mathbf{W_j})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{X} \mid \mathbf{W_C})\, P(\mathbf{W_C}) > P(\mathbf{X} \mid \mathbf{W_j})\, P(\mathbf{W_j})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{W_C}, \mathbf{X})\, P(\mathbf{X}) < P(\mathbf{W_j}, \mathbf{X})\, P(\mathbf{X})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{W_C} \mid \mathbf{X})\, P(\mathbf{X}) < P(\mathbf{W_j} \mid \mathbf{X})\, P(\mathbf{X})$, j=1, 2, ..., N, $j \neq c$

7. Considere um modelo escondido de *Markov* típico para modelação de fonemas: com 3 estados, $q_t \in \{1, 2, 3\}$; a matriz das probabilidades de transição entre estados é (notação Matlab) A=[α 1-α 0; 0 β 1-β; 0 0 1]; admite-se que a cadeia de *Markov* apenas pode iniciar no estado q=1 e apenas pode terminar no estado q=3. Em cada estado está definida uma função densidade de observação, $b_i(\mathbf{x})$, i=1, 2, 3. O valor da probabilidade conjunta P(Q,**X**), em que Q={q_1=1, q_2=2, q_3=2, q_4=3} é um caminho da cadeia de Markov e **X**={$\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$} é uma sequência de vectores, é representada pela expressão:

- □ 1 $b_1(\mathbf{x}_1)$ (1-α) $b_2(\mathbf{x}_2)$ β $b_2(\mathbf{x}_3)$ (1-β) $b_3(\mathbf{x}_4)$.
- □ 1 $b_1(\mathbf{x}_1)$ (1-α) $b_2(\mathbf{x}_2)$ α $b_2(\mathbf{x}_3)$ (1-β) $b_3(\mathbf{x}_4)$.
- □ 1 $b_1(\mathbf{x}_1)$ α $b_2(\mathbf{x}_2)$ β $b_2(\mathbf{x}_3)$ (1-β) $b_3(\mathbf{x}_4)$.
- □ Todas as outras respostas são incorrectas.

8. Dependendo da maneira como é definida a função densidade de probabilidade de observação de um vector de características em cada estado de um *modelo escondido de Markov* (HMM), este pode ser, entre outros, do tipo *semi-contínuo* (SC-HMM) ou *contínuo* (C-HMM). Essa função é geralmente definida combinando linearmente funções gausseanas, as quais são:

- □ partilhadas por vários estados, no caso dos C-HMMs, permitindo o treino mais robusto dos parâmetros.
- □ partilhadas por vários estados, no caso dos SC-HMMs, permitindo o treino mais robusto dos parâmetros.
- □ partilhadas por vários estados, no caso dos C-HMMs, com o objectivo de obter modelos mais precisos.
- □ partilhadas por vários estados, no caso dos SC-HMMs, com o objectivo de obter modelos mais precisos.

9. Em relação às tecnologias dominantes no reconhecimento automático de fala, *modelos escondidos de Markov* (HMM) e *redes neuronais artificiais* (ANN), é correcto afirmar:

- □ Os HMMs apresentam geralmente maior capacidade discriminativa e suportam mais eficazmente o problema da distorção temporal do sinal da fala.
- □ Os ANNs apresentam geralmente maior capacidade discriminativa mas têm mais dificuldade em lidar com o problema da distorção temporal do sinal da fala.
- □ Os sistemas híbridos tentam aliar a capacidade discriminativa dos HMMs com a facilidade de treino dos ANNs.
- □ Os sistemas híbridos utilizam os HMMs para modelar as distribuições de verosimilhança em cada estado e os ANNs para modelar as probabilidades de transição entre os estados.

10. Em relação ao processo de descodificação em sistemas de reconhecimento de fala contínua baseados em modelos escondidos de *Markov* , é correcto afirmar:

- □ O algoritmo *Viterbi* permite identificar a sucessão de palavras correspondente ao alinhamento óptimo entre as duas sucessões de vectores de características.
- □ O algoritmo *forward* permite calcular a verosimilhança acústica mas não identifica a sucessão de palavras correspondente à hipótese de reconhecimento.
- □ O algoritmo *Viterbi* permite identificar a sucessão de palavras correspondente ao “melhor caminho” mas não calcula uma boa aproximação da verosimilhança acústica.
- □ Todas as respostas anteriores são incorrectas.

Teste 6

1. Nos sons vozeados, a frequência fundamental do sinal de fala:
- □ depende directamente do comprimento do tracto vocal.
- □ está relacionada com o comprimento das cordas vocais.
- □ depende da posição dos articuladores na cavidade oral.
- □ Todas as respostas são correctas.

2. Em geral, as vogais caracterizam-se por:
- □ duração curta e pouco variável.
- □ amplitude e taxa de passagem por zero relativamente elevadas.
- □ amplitude relativamente elevada e energia concentrada às altas frequências.
- □ taxa de passagem por zero relativamente baixa e energia concentrada às baixas frequências.

3. O sinal correspondente a consoantes fricativas não vozeadas:
- □ apresenta um forte componente periódica.
- □ apresenta amplitude média relativamente baixa.
- □ apresenta, na sua representação espectral, a energia concentrada às baixas frequências.
- □ Todas as respostas são correctas.

4. Em geral, as consoantes oclusivas apresentam uma fase final:
- □ de silêncio.
- □ durante a qual o som resulta de excitação vozeada.
- □ durante a qual o som resulta de excitação não vozeada.
- □ Todas as respostas anteriores são incorrectas.

5. Se na sequência fonética da palavra “vaga” se negar apenas a característica de vozeamento das consoantes, obter-se-ia a seguinte palavra:
- □ saca.
- □ faca.
- □ fala.
- □ sala.

6. A produção oral da vogal [α] como na palavra “nas” é associada a ressonâncias do tracto vocal. No caso do comprimento efectivo do tracto vocal ter o valor 15 cm a mais baixa frequência de ressonância terá o valor aproximado (considere a velocidade de propagação da onda igual a 343 m/s) de:
- □ 572 Hz.
- □ 857 Hz.
- □ 1143 Hz.
- □ Todas as respostas anteriores são incorrectas.

7. As características mecânicas da membrana basilar são importantes para que o ouvido humano seja capaz de detectar e discriminar frequências diferentes. É correcto afirmar que:
- □ a posição, ao longo da membrana, onde a amplitude de vibração é máxima depende da frequência do sinal acústico.
- □ a percepção auditiva é reforçada fortemente quando dois sons com frequências próximas são escutados em simultâneo.
- □ as chamadas bandas críticas mantêm-se aproximadamente constantes ao longo do espectro de audição.
- □ Todas as respostas são correctas.

8. Em relação ao ouvido humano normal, é correcto afirmar que o limiar de audição (valor mínimo da intensidade da onda acústica para que seja audível):

- □ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 100 Hz.
- □ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 20 KHz.
- □ não depende da frequência do sinal acústico.
- □ Todas as respostas anteriores são incorrectas.

9. A análise do sinal da fala é efectuada sobre segmentos (frames) obtidos através de um processo de "janelamento". Em geral, as janelas utilizadas:

- □ têm duração superior à duração média dos segmentos de quase-estacionaridade do sinal.
- □ têm duração de aproximadamente 10 a 30 ms.
- □ têm forma rectangular, de maneira a tentar reduzir a contaminação de qualquer componente do espectro pelas componentes espectrais mais afastadas.
- □ Todas as respostas anteriores são incorrectas.

10. Na zona de valores elevados de taxa de passagem por zero e valores baixos ou médios de energia média encontram-se sinais relativos a:

- □ sons vozeados.
- □ sons não vozeados.
- □ silêncio com "offset".
- □ Todas as respostas anteriores são incorrectas.

11. Os espectrogramas:

- □ de banda-larga são particularmente indicados para visualizar os harmónicos de excitação no sinal da fala.
- □ de banda-larga não permitem uma boa análise dos formantes.
- □ de banda-estreita mostram claramente os harmónicos de excitação.
- □ de banda-estreita mostram claramente os formantes.

12. Os métodos que geralmente se utilizam para determinar a frequência fundamental de um sinal de fala são:

- □ A amplitude média e a energia média.
- □ A amplitude média e a diferença média de amplitude.
- □ A energia média e a autocorrelação.
- □ Todas as respostas anteriores são incorrectas.

13. O modelo de predição linear, utilizado na análise do sinal da fala:

- □ não está relacionado com o modelo "Excitação-Filtro" do processo de produção da fala.
- □ apresenta limitações importantes na modelação de determinadas classes de sons, como por exemplo as nasais, onde se verificam anti-ressonâncias.
- □ necessita um número muito elevado de parâmetros, tipicamente algumas dezenas, para modelar convenientemente as consoantes fricativas.
- □ Todas as respostas anteriores são incorrectas.

14. Na tecnologia da fala, em geral utiliza-se a análise cepstral para:

- □ determinar apenas a componente de excitação associada ao sinal da fala.
- □ determinar os parâmetros da onda glotal.
- □ poder desconvolucionar o sinal de excitação e a resposta impulsional do filtro correspondente ao tracto vocal.
- □ Todas as respostas anteriores são incorrectas.

15. A ordem mais correcta das operações de processamento linguístico do texto para TTS será:

- □ pré-processamento, análise sintática, análise morfológica, conversão fonética;
- □ análise sintática, análise morfológica, conversão fonética, divisão silábica;
- □ pré-processamento, análise morfológica , análise sintática, conversão fonética;
- □ pré-processamento, análise sintática, conversão fonética , análise morfológica.

16. A acentuação das sílabas em português define-se na presença de acentos gráficos, na última sílaba, se for, por exemplo "az", e na penúltima sílaba, em geral. Isto tem importância para:

- □ controlar a f0 da sílaba;
- □ controlar a duração da sílaba;
- □ controlar a intensidade da sílaba;
- □ todas as respostas são correctas.

17. O grande objectivo da análise prosódica é definir os padrões de entoação (f0) e de durações dos fonemas a sintetizar de forma a comandar o módulo de geração de sinal. O padrão de entoação engloba:

- □ padrão para a frase;
- □ padrão para cada palavra;
- □ uma subida final no caso de uma frase interrogativa;
- □ a segunda resposta e a terceira resposta são correctas.

---

18. Descreva resumidamente os elementos essenciais do processo de produção dos sons da fala, considerando os diferentes tipos de excitação.

19. Explique sucintamente uma técnica que possa ser utilizada para determinar a frequência fundamental do sinal da fala.

20. De uma maneira resumida pode afirmar-se que um conversor texto-fala (TTS) procede à geração de fala correspondente a uma forma convencionada de falar um determinado texto. Explique sucintamente em que consiste o processamento prosódico de um texto dado e quais os seus objectivos.

Teste 7

1. Em geral, o reconhecimento automático de fala:

- □ tem como principal objectivo a identificação do idioma ou dialecto.
- □ permite extrair informação linguística associada ao sinal da fala.
- □ baseia-se no reconhecimento automático do orador.
- □ Todas as respostas anteriores são incorrectas.

2. Considere duas aplicações, A e B, de reconhecimento de fala que se podem distinguir pela perplexidade linguística (Q), pela relação sinal/ruído (SNR) e pela dependência, ou independência, relativamente ao falante. Atendendo apenas a estes três factores de dificuldade, a aplicação A é mais exigente que a aplicação B se:

- □ Q(A) > Q(B) ; SNR(A) > SNR(B) ; A e B são dependentes do falante.
- □ Q(A) < Q(B) ; SNR(A) < SNR(B) ; A e B são independentes do falante.
- □ Q(A) = Q(B) ; SNR(A) > SNR(B) ; A é independente do falante e B é dependente.
- □ Q(A) = Q(B) ; SNR(A) < SNR(B) ; A é independente do falante e B é dependente.

3. Os sistemas de reconhecimento automático de fala apresentam um módulo inicial que extrai as características do sinal acústico. Este módulo de análise:

- □ integra também os algoritmos para "alinhamento temporal", dedicados à solução do problema da distorção temporal do sinal.
- □ deveria, idealmente, extrair apenas a informação relevante para o processo de classificação, favorecendo a separação entre classes.
- □ conduz a uma representação compacta do sinal, baseada em sequências de vectores de características definidos num espaço acústico de pequena dimensão, tipicamente inferior a 10.
- □ Todas as respostas são correctas.

4. O módulo de classificação de muitos sistemas de reconhecimento de fala baseia-se numa estrutura em "rede" que integra a informação relativa ao modelo acústico e ao modelo linguístico. Esta abordagem:

- □ exige um módulo de análise do sinal relativamente complexo, facto que constitui a sua principal desvantagem.
- □ é particularmente eficiente na maneira como permite a interacção dos modelos acústico e linguístico, ambos geralmente definidos com base em regras simbólicas.
- □ não pode ser aplicada quando se utilizam modelos acústicos, correspondentes às unidades linguísticas elementares, definidos ao nível da palavra.
- □ Todas as respostas anteriores são incorrectas.

5. O processo de treino dos modelos acústicos nos sistemas de reconhecimento de fala:

- □ é frequentemente não discriminativo, apesar de exigir maior esforço computacional que o treino discriminativo.
- □ estabelece as fronteiras, definidas no espaço de representação acústica do sinal da fala, entre as diferentes classes linguísticas.
- □ é, em geral, supervisionado por não exigir a prévia segmentação e anotação de dados para treino.
- □ Todas as respostas são correctas.

6. É essencial que os sistemas de reconhecimento automático de fala apresentem uma boa capacidade de generalização, objectivo que se tenta alcançar:

- □ aumentando o mais possível o número de iterações de treino, independentemente da relação entre o número de parâmetros livres do sistema e o tamanho da base de dados para treino.
- □ aumentando o mais possível o número de parâmetros livres do sistema, independentemente do tamanho da base de dados para treino e do número de iterações de treino.
- □ aumentando o mais possível o número de iterações de treino e o número de parâmetros livres, independentemente do tamanho da base de dados para treino.
- □ Todas as respostas anteriores são incorrectas.

7. Seja **X** uma sucessão de vectores de características extraídos de um segmento de sinal de fala que corresponde, por hipótese, à classe **W**. Um sistema de reconhecimento automático de fala utiliza um modelo:

- □ linguístico que calcula o valor de P(**X**|**W**), a probabilidade *a priori*.
- □ linguístico que calcula o valor de P(**W**), a probabilidade *a posteriori*.
- □ acústico que calcula o valor de P(**X**|**W**), a verosimilhança acústica.
- □ acústico que calcula o valor de P(**W**|**X**), a probabilidade *a posteriori*.

8. Seja **X** uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja $\mathbf{W}_j$, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um classificador Bayesiano classifica **X** como pertencente à classe $\mathbf{W_C}$ se e só se:

- □ $P(\mathbf{W_C})\ P(\mathbf{X} \mid \mathbf{W_C}) > P(\mathbf{W_j})\ P(\mathbf{X} \mid \mathbf{W_j})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{X}, \mathbf{W_C}) > P(\mathbf{X}, \mathbf{W_j})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{W_C} \mid \mathbf{X}) > P(\mathbf{W_j} \mid \mathbf{X})$, j=1, 2, ..., N, $j \neq c$
- □ Todas as respostas são correctas.

9. Uma das dificuldades importantes nas tarefas de reconhecimento, até nas mais simples, tem origem no problema da "deformação temporal" típica do sinal da fala. Alguns sistemas baseados na técnica de classificação *template matching* tentam abordar eficazmente este problema utilizando o algoritmo DTW, o qual efectua o alinhamento temporal:

- □ linear de duas sucessões de vectores minimizando uma determinada medida da distância entre elas.
- □ linear de duas sucessões de vectores de maneira a correlacionar os seus comprimentos.
- □ não linear de duas sucessões de vectores de maneira a minimizar a diferença entre os seus comprimentos.
- □ Todas as respostas anteriores são incorrectas.

10. Considere uma *cadeia de Markov* com S+2 estados: $q_t \in \{0, 1, ..., S+1\}$ ($q_t=0$ e $q_t=S+1$ são os valores da variável aleatória correspondentes aos estados "fictícios" inicial e final, respectivamente).

Seja $A=[a_{i,j}]$, i, j=0, 1, ..., S+1, a matriz das probabilidades de transição entre estados, com $a_{i,j}=0$ se $j<i$ ou $j>i+1$. Seja $Q=\{ q_0=0, q_1, q_2, ..., q_T, q_{T+1}=S+1\}$ um *caminho* com comprimento T+2. Então, deve verificar-se:

- □ $T \geq S$ e $P(Q) = \prod_{t=0}^{T} a_{q_t, q_{t+1}}$.
- □ $T \geq S$ e $P(Q) = \prod_{t=1}^{T-1} a_{q_t, q_{t+1}}$.
- □ $T < S$ e $P(Q) = \prod_{t=0}^{T} a_{q_t, q_{t+1}}$.
- □ Todas as outras respostas são incorrectas.

11. Dependendo da maneira como é definida a densidade de probabilidade de observação de um vector de características em cada estado de um *modelo escondido de Markov* (HMM), este pode ser do tipo *discreto* (D-HMM), *semi-contínuo* (SC-HMM) ou *contínuo* (C-HMM). Essa função é geralmente definida combinando linearmente funções elementares, tipicamente gausseanas, que são:

- □ partilhadas com outros estados, no caso dos C-HMMs.
- □ partilhadas com outros estados, no caso dos SC-HMMs.
- □ definidas independentemente em cada estado, no caso dos D-HMMs.
- □ definidas em sub-domínios disjuntos no espaço de representação acústica, no caso dos D-HMMs.

12. O algoritmo *Viterbi* e o algoritmo *forward* permitem o cálculo da verosimilhança acústica em sistemas baseados nos modelos escondidos de Markov. Relativamente ao reconhecimento de fala contínua é correcto afirmar que:

- □ o algoritmo *forward* permite identificar a sucessão de palavras correspondente ao "melhor caminho".
- □ o algoritmo *Viterbi* permite identificar a sucessão de palavras correspondente ao alinhamento óptimo entre as duas sucessões de vectores de características.
- □ o algoritmo *Viterbi* permite identificar a sucessão de palavras correspondente ao "melhor caminho" mas é computacionalmente bastante mais exigente que o algoritmo *forward*.
- □ Todas as respostas anteriores são incorrectas.

13. A modelação acústica em sistemas como o proposto no trabalho prático "Reconhecedor automático de palavras isoladas" pode basear-se nos *modelos escondidos de Markov* (HMM) ou nas *redes neuronais artificiais* (ANN). Neste contexto, comparando estas duas tecnologias, HMM *versus* ANN, é correcto afirmar que os HMMs treinados segundo o critério da máxima verosimilhança apresentam:

- □ maior capacidade discriminativa, pois o critério de treino é discriminativo, e suportam mais facilmente a distorção temporal do sinal da fala.
- □ menor capacidade discriminativa e não suportam tão eficazmente a distorção temporal do sinal da fala.
- □ menor capacidade discriminativa mas suportam mais eficazmente a distorção temporal do sinal da fala.
- □ Todas as respostas anteriores são incorrectas.

14. Em geral, os sistemas de reconhecimento automático de fala contínua com vocabulário de grande dimensão baseiam o modelo acústico global em modelos acústicos elementares definidos a um nível *sub-palavra*. Esta abordagem:

- □ apresenta o inconveniente de exigir mais espaço de memória relativamente a um sistema que utilize, por exemplo, modelos elementares definidos ao nível da palavra.
- □ contribui decisivamente para a utilização mais eficiente dos dados existentes para o treino dos modelos.
- □ conduz a modelos acústicos mais precisos, sobretudo quando estes são treinados sem informação de contexto.
- □ Todas as respostas são correctas.

15. Considere um sistema de reconhecimento de fala contínua, dedicado a uma aplicação com vocabulário de grande dimensão, cujo modelo acústico é baseado em unidades elementares ao nível fonético. Em geral, o modelo linguístico do sistema:

- □ utiliza um léxico para reduzir a *perplexidade* da tarefa de reconhecimento.
- □ utiliza uma gramática baseada em regras *fonéticas*.
- □ assenta numa gramática do tipo N-Gram, tipicamente com o valor de N superior a 2, treinada com o mesmo conjunto de dados utilizados no treino do modelo acústico.
- □ Todas as respostas anteriores são incorrectas.

16. De maneira sucinta, descreva um sistema para reconhecimento de fala contínua e apresente as tarefas mais importantes e os recursos necessários para o seu desenvolvimento.

17. Explique resumidamente o que é a capacidade de generalização de um reconhecedor automático de fala. Indique, justificando, algumas medidas que deverão ser adoptadas, quer na definição da estrutura dos modelos estatísticos quer no seu treino, com o objectivo de melhorar a capacidade de generalização.

18. Em grande parte dos sistemas de reconhecimento automático de fala, o modelo acústico global baseia-se em modelos elementares definidos a um nível sub-palavra. Explique resumidamente quais são os aspectos mais importantes que devem ser considerados na escolha da unidade acústica elementar.

19. Explique de maneira sucinta o que é a capacidade discriminativa dos modelos acústicos elementares em sistemas de reconhecimento automático de fala. Aponte um procedimento, na fase de treino dos modelos, essencial para aumentar a essa capacidade.

20. Descreva de maneira sucinta as etapas mais significativas do processo de desenvolvimento de um sistema, baseado na tecnologia dos modelos escondidos de Markov ou em alternativa nas redes neuronais artificiais, para uma aplicação de reconhecimento de palavras isoladas similar à proposta no trabalho prático “Reconhecedor automático de palavras isoladas”.

Teste 8

1. Seja **X** uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja $\mathbf{W}_j$, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um reconhecedor de fala baseado no critério Bayesiano de classificação reconhece **X** como pertencente à classe $\mathbf{W}_c$ se e só se:

- □ $P(\mathbf{X}, \mathbf{W}_c)\, P(\mathbf{W}_c) > P(\mathbf{X}, \mathbf{W}_j)\, P(\mathbf{W}_j)$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{X} \mid \mathbf{W}_c)\, P(\mathbf{W}_c) > P(\mathbf{X} \mid \mathbf{W}_j)\, P(\mathbf{W}_j)$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{W}_c, \mathbf{X})\, P(\mathbf{X}) < P(\mathbf{W}_j, \mathbf{X})\, P(\mathbf{X})$, j=1, 2, ..., N, $j \neq c$
- □ $P(\mathbf{W}_c \mid \mathbf{X})\, P(\mathbf{X}) < P(\mathbf{W}_j \mid \mathbf{X})\, P(\mathbf{X})$, j=1, 2, ..., N, $j \neq c$

2. Em relação às tecnologias dominantes no reconhecimento automático de fala, *modelos escondidos de Markov* (HMM) e *redes neuronais artificiais* (ANN), é correcto afirmar:

- □ Os HMMs apresentam geralmente maior capacidade discriminativa e suportam mais eficazmente o problema da distorção temporal do sinal da fala.
- □ Os ANNs apresentam geralmente maior capacidade discriminativa mas têm mais dificuldade em lidar com o problema da distorção temporal do sinal da fala.
- □ Os sistemas híbridos tentam aliar a capacidade discriminativa dos HMMs com a facilidade de treino dos ANNs.
- □ Os sistemas híbridos utilizam os HMMs para modelar as distribuições de verosimilhança em cada estado e os ANNs para modelar as probabilidades de transição entre os estados.

3. Seja **X** uma sucessão de vectores de características extraídos de um segmento de sinal de fala que corresponde, por hipótese, à classe **W**. O módulo de Classificação de um sistema de reconhecimento de fala calcula o valor:

- □ da probabilidade *a priori*, P(**X**|**W**), através de um modelo Linguístico.
- □ da verosimilhança acústica, P(**X**|**W**), através de um modelo Acústico.
- □ da probabilidade *a posteriori*, P(**W**), através de um modelo Linguístico.
- □ Todas as respostas anteriores são incorrectas.

4. No sentido de aumentar a capacidade de generalização dos sistemas de reconhecimento automático de fala, em geral:

- □ aumenta-se o mais possível o número de iterações de treino e o número de parâmetros livres, independentemente do tamanho da base de dados para treino.
- □ diminui-se o mais possível o número de parâmetros livres do sistema, independentemente do tamanho da base de dados para treino e do número de iterações de treino.
- □ durante o processo de treino utiliza-se um conjunto de dados, ainda não observados pelo sistema, para acompanhar a evolução da capacidade de generalização e agir em conformidade.
- □ termina-se o processo de treino quando o erro estimado sobre o conjunto de treino é mínimo, desde que o número de parâmetros livres do sistema seja suficientemente elevado.

5. Os sistemas de reconhecimento automático da fala apresentam um módulo inicial que extrai as características do sinal acústico. Este módulo de análise:

- □ Integra os algoritmos para "alinhamento temporal", dedicados à solução do problema da distorção temporal do sinal
- □ Deveria idealmente extrair apenas a informação relevante para o processo de classificação, favorecendo a separação entre classes
- □ Conduz a uma representação compacta do sinal, baseada na sequência de vectores de características definidos num espaço acústico de pequena dimensão, tipicamente inferior a 10
- □ Todas as respostas estão correctas

6. Em geral o reconhecimento automático da fala:

- □ Tem como principal objectivo a identificação do idioma ou dialecto
- □ Permite extrair informação linguística associada ao sinal da fala.
- □ Baseia-se no reconhecimento automático do orador
- □ Todas as respostas anteriores são incorrectas

7. Uma das dificuldades importantes nas tarefas de reconhecimento, até nas mais simples, tem origem no problema da deformação temporal típica do sinal de fala. Alguns sistemas baseados na técnica de classificação "template matching" tentam abordar eficazmente esse problema utilizando o algoritmo DTW, o qual efectua o alinhamento temporal:

- □ Linear de duas sucessões de vectores minimizando uma determinada medida de distância entre eles.
- □ Não linear de duas sucessões de vectores de maneira a minimizar a diferença entre os seus comprimentos.
- □ Linear de duas sucessões de vectores de maneira a correlacionar os seus comprimentos.
- □ Todas as respostas anteriores são incorrectas.

8. O algoritmo Viterbi e o algoritmo "forward" permitem o cálculo da verossimilhança acústica em sistemas baseados nos modelos escondidos de Markov. Relativamente ao reconhecimento da fala contínua é correcto afirmar que:

- □ O algoritmo forward permite identificar a sucessão de palavras correspondente ao "melhor caminho"
- □ O algoritmo Viterbi permite identificar a sucessão de palavras correspondente ao "melhor caminho" mas é computacionalmente mais exigente que o algoritmo forward
- □ O algoritmo Viterbi permite identificar a sucessão de palavras correspondente ao alinhamento óptimo entre as duas sucessões de vectores de características
- □ Todas as respostas anteriores são incorrectas

9. Considere um sistema de reconhecimento de fala contínua, dedicado a uma aplicação com vocabulário de grande dimensão, cujo modelo acústico é baseado em unidades elementares ao nível fonético. Em geral, o modelo linguístico do sistema:

- □ Utiliza uma gramática baseada em regras fonéticas
- □ Utiliza um léxico para reduzir a perplexidade da tarefa de reconhecimento
- □ Assenta numa gramática do tipo n-gram, tipicamente com o valor de n superior a 2, treinada com o mesmo conjunto de dados utilizado no treino do modelo acústico
- □ Todas as respostas anteriores são incorrectas

10. Nos sistemas de reconhecimento de fala , o módulo de Análise:

- □ deveria, idealmente, extrair apenas a informação discriminante para a tarefa de reconhecimento.
- □ transforma, em geral, o sinal acústico de fala numa sequência de vectores de características.
- □ deve conduzir a uma representação onde a variabilidade em cada classe é relativamente pequena e a separação entre classes é relativamente grande.
- □ Todas as respostas são correctas.

11. Explique resumidamente o que é a capacidade de generalização de um reconhecedor automático de fala. Indique, justificando, algumas medidas que deverão ser adoptadas, quer na definição da estrutura dos modelos estatísticos quer no seu treino, com o objectivo de melhorar a capacidade de generalização.

12. Em grande parte dos sistemas de reconhecimento automático de fala, o modelo acústico global baseia-se em modelos elementares definidos a um nível sub-palavra. Explique resumidamente quais são os aspectos mais importantes que devem ser considerados na escolha da unidade acústica elementar.

Teste 9

1. Considere a palavra “javali”. Se na sequência fonética trocarmos apenas a característica de vozeamento das 2 primeiras consoantes qual seria a sequência textual que se obteria:

a) xabali.
b) zavali.
c) fazali.
d) xafali.

2. A produção oral das vogais é associada a ressonâncias do tracto vocal. As respectivas frequências denominam-se formantes. Quanto a estas pode afirmar-se o seguinte:

a) São harmónicos.
b) Dependem da articulação.
c) Dependem do tom.
d) Não dependem da velocidade do som no ar .

3. No modelo fonte-filtro de produção de fala há lugar:

a) Ao sinal excitador com pulsos, ruído ou mistura dos dois tipos.
b) Á função de transferência do trato vocal.
c) À correcção da radiação da boca.
d) todas as alíneas estão certas.

4. Numa zona de sinal com valores muito reduzidos de taxa de passagens por zero pode encontrar-se:

a) silêncio com “offset”.
b) Sinais vozeados.
c) silêncio.
d) Não vozeados.

5. Na determinação de f0 de um segmento de sinal de voz utilizou-se a técnica da autocorrelação e detecção de picos. Para tal deveria ter-se utilizado:

a) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual.
b) Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo menor que o menor período a medir.
c) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual ao período a medir.
d) Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo igual.

6. O espectrograma de um sinal de voz é um exemplo de análise deslizante de termo curto. Utiliza-se uma janela temporal e o resultado da transformada de cada segmento mostra o espectro de termo curto. Entre as variáveis livres para ajustar a visualização das características do sinal pode utilizar-se a escolha do tipo de janela, a sua duração e a taxa de sobreposição:

a) A duração da janela temporal utilizada influencia fortemente a resolução frequencial obtida sendo melhor a resolução obtida com janela mais extensa.
b) O passo a utilizar influencia a resolução frequencial obtida.
c) A duração da janela temporal utilizada não influencia sensivelmente a resolução frequencial obtida.
d) A visualização dos formantes faz-se de preferência com janelas de mais longa duração.

7. No sinal de voz a prosódia consiste:

a) Na entoação.
b) Alínea a) mais padrão de durações segmentais.
c) Alínea b) mais relação sinal/ruído .
d) Alínea c) mais energia.

8. Sabendo que as frequências formantes do sinal de fala indicam certas características do tracto vocal que o produziu, podemos determinar essas frequências:

a) Através da determinação dos picos do espectro do sinal de voz determinados com janela de comprimento adequado.
b) De forma directa da amplitude do sinal de voz.
c) Por meio dos coeficientes LPC do sinal de excitação glotal.
d) Todas as outras alíneas estão erradas.

9. A análise ou codificação LPC consiste na determinação de um conjunto de N coeficientes de uma função descritiva do sistema de produção da fala:

a) Os coeficientes são as raízes da função do sistema.
b) As raízes do sistema de equações normais são os coeficientes.
c) As N+1 equações normais são simétricas.
d) Para construir o sistema de equações normais é necessário calcular (N+1)*N valores diferentes.

10. A frequência fundamental de um sinal de fala pode calcular-se através de:

a) autocorrelação, AMDF, picos e vales e coeficientes LPC
b) Os da alínea a) e a covariância
c) Detecção de máximos ou mínimos de uma certa função calculada a partir do sinal de interesse
d) Todas as outras alíneas estão erradas

11. O pré-processamento do texto contribui para reduzir várias deficiências do texto e realizar conversões. Dê alguns exemplos significativos e ilustrativos dessas funções e discuta os objectivos.

12. Explique o funcionamento glotal durante a produção da fala vozeada.

Teste 10

1. Considere a palavra “vigas”. Se na sequência fonética invertermos apenas a característica de vozeamento de cada uma das 2 primeiras consoantes qual seria a sequência textual que se obteria?

2. Explique o conceito de “pulso glotal” e compare-o com o sinal de erro da modelização LPC de sinais de fala.

3. Explique sucintamente uma técnica não baseada na correlação que possa ser utilizada para determinar a frequência fundamental do sinal da fala.

4. A produção oral das vogais é associada a ressonâncias do tracto vocal. As respectivas frequências denominam-se formantes. Quanto a estas pode afirmar-se o seguinte:

a) São harmónicos.

b) Dependem da articulação.

c) Dependem do tom.

d) Não dependem da velocidade do som no ar .

5. No modelo fonte-filtro de produção de fala há lugar:

a) Ao sinal excitador com pulsos, ruído ou mistura dos dois tipos.

b) Á função de transferência do trato vocal.

c) À correcção da radiação da boca.

d) todas as alíneas estão certas.

6. Em relação ao ouvido humano normal, é correcto afirmar que o limiar de audição (valor mínimo da intensidade da onda acústica para que seja audível):

□ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 100 Hz.

□ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 20 KHz.

□ não depende da frequência do sinal acústico.

□ Todas as respostas anteriores são incorrectas.

7. Na determinação de f0 de um segmento de sinal de voz utilizou-se a técnica da autocorrelação e detecção de picos. Para tal deveria ter-se utilizado:

□ Uma janela de duração bastante superior ao valor médio do período a medir e passo igual.

□ Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo menor que o menor período a medir.

□ Uma janela de duração bastante superior ao valor médio do período a medir e passo igual ao período a medir.

□ Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo igual.

8. O espectrograma de um sinal de voz é um exemplo de análise deslizante de termo curto. Utiliza-se uma janela temporal e o resultado da transformada de cada segmento mostra o espectro de termo curto. Entre as variáveis livres para ajustar a visualização das características do sinal pode utilizar-se a escolha do tipo de janela, a sua duração e a taxa de sobreposição:

□ A duração da janela temporal utilizada influencia fortemente a resolução frequencial obtida sendo melhor a resolução obtida com janela mais extensa.

□ O passo a utilizar influencia a resolução frequencial obtida.

□ A duração da janela temporal utilizada não influencia sensivelmente a resolução frequencial obtida.

□ A visualização dos formantes faz-se de preferência com janelas de mais longa duração.

9. A análise ou codificação LPC consiste na determinação de um conjunto de N coeficientes de uma função descritiva do sistema de produção da fala:

□ Os coeficientes são as raízes da função do sistema.

□ As raízes do sistema de equações normais são os coeficientes.

□ As N+1 equações normais são simétricas.

□ Para construir o sistema de equações normais é necessário calcular (N+1)*N valores diferentes.

10. Na tecnologia da fala, em geral utiliza-se a análise cepstral para:

□ determinar apenas a componente de excitação associada ao sinal da fala.

□ determinar os parâmetros da onda glotal.

□ poder desconvolucionar o sinal de excitação e a resposta impulsional do filtro correspondente ao tracto vocal.

□ Todas as respostas anteriores são incorrectas.

*****

Teste 11

1. Considere a palavra “vigas”. Se na sequência fonética invertermos apenas a característica de vozeamento de cada uma das 2 primeiras consoantes qual seria a sequência textual que se obteria?

2. Explique o conceito de “pulso glotal” e compare-o com o sinal de erro da modelização LPC de sinais de fala.

3. Explique sucintamente uma técnica não baseada na correlação que possa ser utilizada para determinar a frequência fundamental do sinal da fala.

4. A produção oral das vogais é associada a ressonâncias do tracto vocal. As respectivas frequências denominam-se formantes. Quanto a estas pode afirmar-se o seguinte:

□ São harmónicos.

□ Dependem da articulação.

□ Dependem do tom.

□ Não dependem da velocidade do som no ar .

5. No modelo fonte-filtro de produção de fala há lugar:

□ Ao sinal excitador com pulsos, ruído ou mistura dos dois tipos.

□ Á função de transferência do trato vocal.

□ À correcção da radiação da boca.

□ todas as alíneas estão certas.

6. Em relação ao ouvido humano normal, é correcto afirmar que o limiar de audição (valor mínimo da intensidade da onda acústica para que seja audível):

□ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 100 Hz.

□ tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 20 KHz.

□ não depende da frequência do sinal acústico.

□ Todas as respostas anteriores são incorrectas.

7. Na determinação de f0 de um segmento de sinal de voz utilizou-se a técnica da autocorrelação e detecção de picos. Para tal deveria ter-se utilizado:

□ Uma janela de duração bastante superior ao valor médio do período a medir e passo igual.

□ Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo menor que o menor período a medir.

□ Uma janela de duração bastante superior ao valor médio do período a medir e passo igual ao período a medir.

□ Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo igual.

8. O espectrograma de um sinal de voz é um exemplo de análise deslizante de termo curto. Utiliza-se uma janela temporal e o resultado da transformada de cada segmento mostra o espectro de termo curto. Entre as variáveis livres para ajustar a visualização das características do sinal pode utilizar-se a escolha do tipo de janela, a sua duração e a taxa de sobreposição:

□ A duração da janela temporal utilizada influencia fortemente a resolução frequencial obtida sendo melhor a resolução obtida com janela mais extensa.

□ O passo a utilizar influencia a resolução frequencial obtida.

□ A duração da janela temporal utilizada não influencia sensivelmente a resolução frequencial obtida.

□ A visualização dos formantes faz-se de preferência com janelas de mais longa duração.

9. A análise ou codificação LPC consiste na determinação de um conjunto de N coeficientes de uma função descritiva do sistema de produção da fala:

□ Os coeficientes são as raízes da função do sistema.

□ As raízes do sistema de equações normais são os coeficientes.

□ As N+1 equações normais são simétricas.

□ Para construir o sistema de equações normais é necessário calcular (N+1)*N valores diferentes.

10. Na tecnologia da fala, em geral utiliza-se a análise cepstral para:

□ determinar apenas a componente de excitação associada ao sinal da fala.

□ determinar os parâmetros da onda glotal.

□ poder desconvolucionar o sinal de excitação e a resposta impulsional do filtro correspondente ao tracto vocal.

□ Todas as respostas anteriores são incorrectas.

*****

Teste 12

1. É sabido que as ressonâncias do tracto vocal formam o timbre de muitos dos sons da fala. Mostre como se calcularíam os valores das frequências dos formantes da vogal 6 (SAMPA) (ex.: mesa).

2. Explique o conceito de “tracto vocal” e compare a respectiva função de transferência acústica com o polinómio denominador do filtro IIR da modelização LPC de sinais de fala.

3. Compare sucintamente a técnica da diferença média de amplitude com a técnica de de autocorrelação para determinar a frequência fundamental do sinal da fala.

4. Em relação ao ouvido humano normal, é correcto afirmar que o limiar de audição (valor mínimo da intensidade da onda acústica para que seja audível):

a) tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 100 Hz.

b) tem um valor mínimo quando a onda (tom puro) tem frequência com o valor aproximado de 20 KHz.

c) não depende da frequência do sinal acústico.

d) todas as respostas anteriores são incorrectas.

5. No modelo fonte-filtro de produção de fala há lugar:

a) ao sinal excitador com pulsos, ruído ou mistura dos dois tipos.

b) à função de transferência do trato vocal.

c) à correcção da radiação da boca.

d) todas as alíneas estão certas.

6. Na determinação de f0 de um segmento de sinal de voz utilizou-se a técnica da autocorrelação e detecção de picos. Para tal deveria ter-se utilizado:

a) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual.

b) Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo menor que o menor período a medir.

c) Uma janela de duração bastante superior ao valor médio do período a medir e passo igual ao período a medir.

d) Uma janela de duração ligeiramente superior ao valor médio do período a medir e passo igual.

7. A produção oral das vogais é associada a ressonâncias do tracto vocal. As respectivas frequências denominam-se formantes. Quanto a estas pode afirmar-se o seguinte:

a) São harmónicos.

b) Dependem da articulação.

c) Dependem do tom.

d) Não dependem da velocidade do som no ar .

8. O espectrograma de um sinal de voz é um exemplo de análise deslizante de termo curto. Utiliza-se uma janela temporal e o resultado da transformada de cada segmento mostra o espectro de termo curto. Entre as variáveis livres para ajustar a visualização das características do sinal pode utilizar-se a escolha do tipo de janela, a sua duração e a taxa de sobreposição:

a) A duração da janela temporal utilizada influencia fortemente a resolução frequencial obtida sendo melhor a resolução obtida com janela mais extensa.

b) O passo a utilizar influencia a resolução frequencial obtida.

c) A duração da janela temporal utilizada não influencia sensivelmente a resolução frequencial obtida.

d) A visualização dos formantes faz-se de preferência com janelas de mais longa duração.

9. Na tecnologia da fala, em geral utiliza-se a análise cepstral para:

a) determinar apenas a componente de excitação associada ao sinal da fala.

b) determinar os parâmetros da onda glotal.

c) poder desconvolucionar o sinal de excitação e a resposta impulsional do filtro correspondente ao tracto vocal.

d) Todas as respostas anteriores são incorrectas.

10. A análise ou codificação LPC consiste na determinação de um conjunto de N coeficientes de uma função descritiva do sistema de produção da fala:

a) Os coeficientes são as raízes da função do sistema.

b) As raízes do sistema de equações normais são os coeficientes.

c) As N+1 equações normais são simétricas.

d) Para construir o sistema de equações normais é necessário calcular (N+1)*N valores diferentes.

*****

Teste 13

1. Descreva um equalizador gráfico de terços de oitava e um equalizador paramétrico. Distinga os dois tipos.

2. Na apresentação dos microfones sem fios realizada na aula teórica, abordou-se o problema da eventual instabilidade que pode ocorrer na ligação RF entre o emissor e uma antena do receptor devido às reflexões e apontou-se uma solução. Descreva essa solução

3. Descreva um sistema de conversão texto-fala que recebe texto ASCII e produz um sinal digital em formato .WAV, baseando-se num diagrama de blocos adequado e explicando, com exemplificação as operações necessárias contidas nesse diagrama, mencione também as tarefas e recursos de base eventualmente necessários para a operação efectiva do sistema além dos algoritmos que estarão compreendidos no diagrama.

4. Seja X uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja Wj, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um reconhecedor de fala baseado no critério Bayesiano de classificação reconhece X como pertencente à classe WC se e só se:

a) $P(X , WC)\ P(WC) > P(X , Wj)\ P(Wj)$, j=1, 2, ..., N, j□c

b) $P(X \mid WC)\ P(WC) > P(X \mid Wj)\ P(Wj)$, j=1, 2, ..., N, j□c

c) $P(WC, X)\ P(X) < P(Wj, X)\ P(X)$, j=1, 2, ..., N, j□c

d) $P(WC \mid X)\ P(X) < P(Wj \mid X)\ P(X)$, j=1, 2, ..., N, j□c

5. Considere um sistema de reconhecimento de fala contínua, dedicado a uma aplicação com vocabulário de grande dimensão, cujo modelo acústico é baseado em unidades elementares ao nível fonético. Em geral, o modelo linguístico do sistema:

a) utiliza um léxico para reduzir a perplexidade da tarefa de reconhecimento.

b) utiliza uma gramática baseada em regras fonéticas.

c) assenta numa gramática do tipo N-Gram, tipicamente com o valor de N superior a 2, treinada com o mesmo conjunto de dados utilizados no treino do modelo acústico.

d) Todas as respostas anteriores são incorrectas.

6. Em geral, os sistemas de reconhecimento automático de fala contínua com vocabulário de grande dimensão baseiam o modelo acústico global em modelos acústicos elementares definidos a um nível sub-palavra. Esta abordagem:

a) apresenta o inconveniente de exigir mais espaço de memória relativamente a um sistema que utilize, por exemplo, modelos elementares definidos ao nível da palavra.

b) contribui decisivamente para a utilização mais eficiente dos dados existentes para o treino dos modelos.

c) conduz a modelos acústicos mais precisos, sobretudo quando estes são treinados sem informação de contexto.

d) todas as respostas são correctas.

7. A modelação acústica em sistemas como o proposto no trabalho prático “Reconhecedor automático de palavras isoladas” pode basear-se nos modelos escondidos de Markov (HMM) ou nas redes neuronais artificiais (ANN). Neste contexto, comparando estas duas tecnologias, HMM versus ANN, é correcto afirmar que os HMMs treinados segundo o critério da máxima verosimilhança apresentam:

a) maior capacidade discriminativa, pois o critério de treino é discriminativo, e suportam mais facilmente a distorção temporal do sinal da fala.

b) menor capacidade discriminativa e não suportam tão eficazmente a distorção temporal do sinal da fala.

c) menor capacidade discriminativa mas suportam mais eficazmente a distorção temporal do sinal da fala.

d) Todas as respostas anteriores são incorrectas.

8. A ordem mais correcta das operações de processamento linguístico do texto para TTS será:

a) pré-processamento, análise sintática, análise morfológica, conversão fonética;

b) análise sintática, análise morfológica, conversão fonética, divisão silábica;

c) pré-processamento, análise morfológica , análise sintática, conversão fonética;

d) pré-processamento, análise sintática, conversão fonética , análise morfológica.

9. A acentuação das sílabas em português define-se na presença de acentos gráficos, na última sílaba, se for, por exemplo "az", e na penúltima sílaba, em geral. Isto tem importância para:

a) controlar a f0 da sílaba;

b) controlar a duração da sílaba;

c) controlar a intensidade da sílaba;

d) todas as respostas são correctas.

10. O grande objectivo da análise prosódica é definir os padrões de entoação (f0) e de durações dos fonemas a sintetizar de forma a comandar o módulo de geração de sinal. O padrão de entoação engloba:

a) padrão para a frase;

b) padrão para cada palavra;

c) uma subida final no caso de uma frase interrogativa;

d) a segunda resposta e a terceira resposta são correctas.

*****

Teste 14

1. Descreva um equalizador gráfico de oitavas e um equalizador paramétrico. Distinga os dois tipos.

2. Considere a função de transferência relacionada com a cabeça (HRTF) e explique a este respeito três aspectos:o seu significado, a sua potencial utilidade e uma possível forma da sua obtenção.

3. Descreva um sistema de conversão texto-fala que recebe texto ASCII e produz um sinal digital em formato .WAV, baseando-se num diagrama de blocos adequado e explicando, com exemplificação as operações necessárias contidas nesse diagrama, mencione também as tarefas e recursos de base eventualmente necessários para a operação efectiva do sistema além dos algoritmos que estarão compreendidos no diagrama.

4. Considere um sistema de reconhecimento de fala contínua, dedicado a uma aplicação com vocabulário de grande dimensão, cujo modelo acústico é baseado em unidades elementares ao nível fonético. Em geral, o modelo linguístico do sistema:

a) utiliza um léxico para reduzir a perplexidade da tarefa de reconhecimento.

b) utiliza uma gramática baseada em regras fonéticas.

c) assenta numa gramática do tipo N-Gram, tipicamente com o valor de N superior a 2, treinada com o mesmo conjunto de dados utilizados no treino do modelo acústico.

d) Todas as respostas anteriores são incorrectas.

5. Em geral, os sistemas de reconhecimento automático de fala contínua com vocabulário de grande dimensão baseiam o modelo acústico global em modelos acústicos elementares definidos a um nível sub-palavra. Esta abordagem:

a) apresenta o inconveniente de exigir mais espaço de memória relativamente a um sistema que utilize, por exemplo, modelos elementares definidos ao nível da palavra.

b) contribui decisivamente para a utilização mais eficiente dos dados existentes para o treino dos modelos.

c) conduz a modelos acústicos mais precisos, sobretudo quando estes são treinados sem informação de contexto.

d) todas as respostas são correctas.

6. Seja X uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja Wj, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um classificador Bayesiano classifica X como pertencente à classe WC

se e só se::

a) $P(WC)\ P(X \mid WC) > P(Wj)\ P(X \mid Wj)$, j=1, 2, ..., N, j□c

b) $P(X , WC) > P(X , Wj)$, j=1, 2, ..., N, j□c

c) $P(WC \mid X) > P(Wj \mid X)$, j=1, 2, ..., N, j□c

d) Todas as respostas são correctas

7. A modelação acústica em sistemas como o proposto no trabalho prático “Reconhecedor automático de palavras isoladas” pode basear-se nos modelos escondidos de Markov (HMM) ou nas redes neuronais artificiais (ANN). Neste contexto, comparando estas duas tecnologias, HMM versus ANN, é correcto afirmar que os HMMs treinados segundo o critério da máxima verosimilhança apresentam:

a) maior capacidade discriminativa, pois o critério de treino é discriminativo, e suportam mais facilmente a distorção temporal do sinal da fala.

b) menor capacidade discriminativa e não suportam tão eficazmente a distorção temporal do sinal da fala.

c) menor capacidade discriminativa mas suportam mais eficazmente a distorção temporal do sinal da fala.

d) Todas as respostas anteriores são incorrectas.

8. O grande objectivo da análise prosódica é definir os padrões de entoação (f0) e de durações dos fonemas a sintetizar de forma a comandar o módulo de geração de sinal. O padrão de entoação engloba:

a) padrão para a frase;

b) padrão para cada palavra;

c) uma subida final no caso de uma frase interrogativa;

d) a segunda resposta e a terceira resposta são correctas.

9. A ordem mais correcta das operações de processamento linguístico do texto para TTS será:

a) pré-processamento, análise sintática, análise morfológica, conversão fonética;

b) análise sintática, análise morfológica, conversão fonética, divisão silábica;

c) pré-processamento, análise morfológica , análise sintática, conversão fonética;

d) pré-processamento, análise sintática, conversão fonética , análise morfológica.

10. A acentuação das sílabas em português define-se na presença de acentos gráficos, na última sílaba, se for, por exemplo "az", e na penúltima sílaba, em geral. Isto tem importância para:

a) controlar a f0 da sílaba;

b) controlar a duração da sílaba;

c) controlar a intensidade da sílaba;

d) todas as respostas são correctas.

*****

Teste 15

1. Descreva detalhadamente um processo para avaliar objectivamente a intelibilidade do campo acústico de uma sala.

2. Considere a função de transferência relacionada com a cabeça (HRTF). Analise e explique qual será o principal problema na audição do material gravado de forma binaural quando fôr realizada num sistema estéreo.

3. Descreva um sistema de conversão texto-fala que recebe texto ASCII e produz um sinal digital em formato .WAV, baseando-se num diagrama de blocos adequado e explicando, com exemplificação as operações necessárias contidas nesse diagrama, mencione também as tarefas e recursos de base eventualmente necessários para a operação efectiva do sistema além dos algoritmos que estarão compreendidos no diagrama.

4. Considere um sistema de reconhecimento de fala contínua, dedicado a uma aplicação com vocabulário de grande dimensão, cujo modelo acústico é baseado em unidades elementares ao nível fonético. Em geral, o modelo linguístico do sistema:

a) utiliza um léxico para reduzir a perplexidade da tarefa de reconhecimento.

b) utiliza uma gramática baseada em regras fonéticas.

c) assenta numa gramática do tipo N-Gram, tipicamente com o valor de N superior a 2, treinada com o mesmo conjunto de dados utilizados no treino do modelo acústico.

d) Todas as respostas anteriores são incorrectas.

5. Em geral, os sistemas de reconhecimento automático de fala contínua com vocabulário de grande dimensão baseiam o modelo acústico global em modelos acústicos elementares definidos a um nível sub-palavra. Esta abordagem:

a) apresenta o inconveniente de exigir mais espaço de memória relativamente a um sistema que utilize, por exemplo, modelos elementares definidos ao nível da palavra.

b) contribui decisivamente para a utilização mais eficiente dos dados existentes para o treino dos modelos.

c) conduz a modelos acústicos mais precisos, sobretudo quando estes são treinados sem informação de contexto.

d) todas as respostas são correctas.

6. Seja X uma sucessão conhecida de vectores de características extraídos de um segmento do sinal de fala e seja Wj, (j=1, 2, ..., N) uma sucessão de símbolos linguísticos correspondentes a uma das N hipóteses de reconhecimento. Um classificador Bayesiano classifica X como pertencente à classe WC

se e só se::

a) $P(WC)\ P(X \mid WC) > P(Wj)\ P(X \mid Wj)$, j=1, 2, ..., N, j□c

b) $P(X , WC) > P(X , Wj)$, j=1, 2, ..., N, j□c

c) $P(WC \mid X) > P(Wj \mid X)$, j=1, 2, ..., N, j□c

d) Todas as respostas são correctas

7. A modelação acústica em sistemas como o proposto no trabalho prático “Reconhecedor automático de palavras isoladas” pode basear-se nos modelos escondidos de Markov (HMM) ou nas redes neuronais artificiais (ANN). Neste contexto, comparando estas duas tecnologias, HMM versus ANN, é correcto afirmar que os HMMs treinados segundo o critério da máxima verosimilhança apresentam:

a) maior capacidade discriminativa, pois o critério de treino é discriminativo, e suportam mais facilmente a distorção temporal do sinal da fala.

b) menor capacidade discriminativa e não suportam tão eficazmente a distorção temporal do sinal da fala.

c) menor capacidade discriminativa mas suportam mais eficazmente a distorção temporal do sinal da fala.

d) Todas as respostas anteriores são incorrectas.

8. O grande objectivo da análise prosódica é definir os padrões de entoação (f0) e de durações dos fonemas a sintetizar de forma a comandar o módulo de geração de sinal. O padrão de entoação engloba:

a) padrão para a frase;

b) padrão para cada palavra;

c) uma subida final no caso de uma frase interrogativa;

d) a segunda resposta e a terceira resposta são correctas.

9. A ordem mais correcta das operações de processamento linguístico do texto para TTS será:

a) pré-processamento, análise sintática, análise morfológica, conversão fonética;

b) análise sintática, análise morfológica, conversão fonética, divisão silábica;

c) pré-processamento, análise morfológica , análise sintática, conversão fonética;

d) pré-processamento, análise sintática, conversão fonética , análise morfológica.

10. A acentuação das sílabas em português define-se na presença de acentos gráficos, na última sílaba, se for, por exemplo "az", e na penúltima sílaba, em geral. Isto tem importância para:

a) controlar a f0 da sílaba;

b) controlar a duração da sílaba;

c) controlar a intensidade da sílaba;

d) todas as respostas são correctas.

*****