

O CONHECIMENTO ARQUIVÍSTICO APLICADO NA DESCRIÇÃO DE PADRÕES FONÉTICOS

EL CONOCIMIENTO ARCHIVÍSTICO APLICADO EN LA DESCRIPCIÓN DE ESTÁNDARES FONÉTICOS

Tauana dos Santos Barbosa*
Jussara Borges**
Antonio César Morant Braid***

RESUMO:

Introdução: A identificação de falantes é a atividade capaz de determinar se duas vozes foram produzidas por uma mesma pessoa. O exame é realizado pelo foneticista, por meio da comparação de padrões fonéticos extraídos de falantes, que são elementos que determinam traços característicos da fala de um indivíduo. **Objetivo:** Este estudo objetiva demonstrar como o conhecimento arquivístico pode ser utilizado para descrever os padrões fonéticos fornecidos pelo foneticista, auxiliando na construção de sistemas de identificação de falantes. **Metodologia:** A pesquisa caracteriza-se como descritiva e de natureza quali-quantitativa, com a interpretação, descrição, avaliação e indexação dos padrões fonéticos, além da aplicação de conceitos de estatística. **Resultado:** O resultado da investigação levou à proposta de um modelo de descrição dos padrões fonéticos, organizado em forma de matriz, com informações relevantes para o foneticista, para ser utilizado em sistemas de identificação de falantes. **Conclusão:** Conclui-se que a prática da interdisciplinaridade entre as duas áreas, Arquivologia e Fonética, pode contribuir de maneira importante nos resultados de organização de dados e informações de natureza fonética, o que demonstra a ampla abrangência do conhecimento arquivísticos na resolução de problemas hodiernos.

Palavras-chave: Descrição arquivística. Gestão da informação. Fonética Forense

*Arquivista pela Universidade Federal da Bahia. E-mail tbarbosa542@gmail.com

**Doutora em Comunicação e Cultura Contemporânea. Professora Adjunta do Instituto de Ciência da Informação da Universidade Federal da Bahia – UFBA. E-mail jussarab@ufba.br

***Perito Criminal do Departamento de Polícia Técnica. E-mail cesarbraid@hotmail.com

1 INTRODUÇÃO

O conhecimento arquivístico pode ter aplicação em diversas áreas de atividades que têm como objeto o documento e sua organização. Por isso, a Arquivologia tem se dedicado a desenvolver conhecimento aplicado à gestão documental, o que envolve sua representação em uma linguagem capaz de recuperação conforme a necessidade do usuário.

Uma área que ainda se ressentida de gestão documental é a Fonética Forense. Atualmente não existe nenhum banco de dados de padrão de vozes no Brasil e, segundo Braid¹, após a coleta das vozes, o material é arquivado da forma como é coletado, em anotações e mídias, sem a devida descrição que possa auxiliar na recuperação da informação. Um dos objetivos desta pesquisa, portanto, foi desenvolver um modelo de descrição dos padrões fonéticos, com todas as informações de relevância para o foneticista, em formato de matriz, baseado nos conhecimentos de descrição arquivística.

Assim, este trabalho pretende aplicar o conhecimento arquivístico na Fonética Forense, auxiliando na descrição das informações contidas nos padrões fonéticos, visando à construção de sistemas de identificação de pessoas pela voz. A utilização da identificação de falantes na área forense é necessária e quase sempre insubstituível na solução de crimes em relação aos quais existem vozes registradas em algum tipo de mídia. Trata-se de atividade capaz de determinar se duas vozes foram produzidas por um mesmo falante, servindo como meio de prova para atribuir a alguém a autoria de um delito, ou desvincular o envolvimento de um inocente a um crime que possa lhe estar sendo imputado.

Os padrões fonéticos serão descritos segundo os conhecimentos arquivísticos, baseando-se, resumidamente, em aspectos de descrição, avaliação, indexação e aplicação da tabela de temporalidade, para se obter maior eficiência no processo de recuperação da informação. Com isso, pretende-se criar um modelo de descrição dos padrões fonéticos, organizados

¹ Informação verbal: informação fornecida pelo foneticista César Braid em janeiro de 2015, que depois viria a ser convidado a contribuir como coautor deste artigo por seu conhecimento e experiência em Fonética Forense.

em forma de matriz, com informações relevantes para o foneticista, que é o profissional responsável pela coleta dos padrões fonéticos.

O artigo organiza-se em três partes principais que se seguem a esta Introdução: na seção seguinte expõe-se brevemente o conceito da Fonética Forense e a geração de documentos no seu âmbito para, em seguida, explicar-se a aplicação da descrição arquivísticas aos documentos gerados; a seção seguinte apresenta um modelo de descrição para padrões fonéticos e, por fim, expõe-se as principais conclusões a que esta investigação chegou.

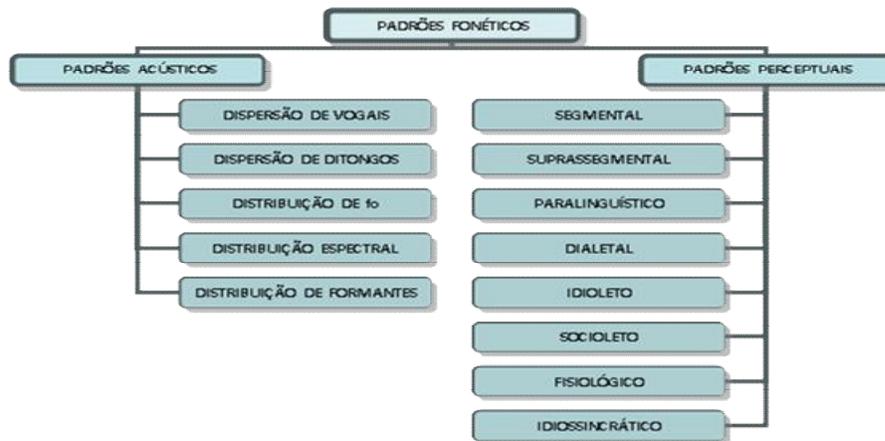
2 FONÉTICA FORENSE

A Fonética (do grego *phonetiké*) tem por objeto a fala, mediante o estudo da produção e percepção dos seus sons (BRAID, 2003, p.5) e forense refere-se aos fatos relacionados à justiça. A Fonética Forense é o estudo da fala através da identificação de falantes que passam por um exame de autoria de voz.

A identificação de pessoas pela voz se baseia na comparação de padrões fonéticos, que são características que são extraídas de vozes gravadas. Para a comparação dos padrões fonéticos por meio de sistemas informatizados é necessário descrevê-los de forma adequada que propiciem a recuperação da informação. Nesse sentido, a Arquivologia pode contribuir propondo modelos de descrição dos padrões fonéticos coletados pelo foneticista.

Os padrões são classificados, conforme Braid (2003), em dois grandes grupos: perceptuais e acústicos (Figura 1). Cada padrão do grupo perceptual, assim como acústico, é sub-classificado de acordo com sua especificação fonética. Por exemplo, os padrões fonéticos dialetais podem ser sub-classificados de acordo com a zona dialetal do falante: dialeto nordestino, baiano, do litoral etc.

Figura 1 - Esquema representativo de padrões fonéticos acústicos e perceptuais



Fonte: Braid, 2003

Observe-se que a voz humana é o resultado acústico de configurações fisiológicas do aparelho de fala, que se altera com a idade. Por essa razão, a voz apresenta modificações ao longo da vida, que devem ser consideradas para determinar a adequabilidade da comparação entre dois padrões. Nessa perspectiva, é importante a elaboração de uma tabela de temporalidade em que conste a adequabilidade de comparação entre dois padrões fonéticos, indicando os arquivos correntes, intermediários e permanentes.

Para uma compreensão da função desempenhada pelo arquivista e pelo foneticista, será abordado aqui, de forma breve, como o trabalho conjunto dos dois profissionais contribui na descrição, sendo que o arquivista é o principal mediador e responsável pela gestão da informação em seus diversos suportes.

O foneticista é responsável pela coleta e extração dos dados contidos nos padrões fonéticos, pois ele é quem detém o conhecimento científico e técnico para tal atividade. Essa é a primeira etapa do trabalho: a coleta e extração dos dados que correspondem a um conjunto de características do falante, formando um padrão fonético que pertence a um determinado indivíduo. Esses dados são fornecidos ao arquivista, que é o responsável pela interpretação, avaliação e criação da descrição com informações de interesse do foneticista para o processo de comparação no sistema de identificação de falantes. O arquivista transforma os dados existentes nos padrões fonéticos em informações precisas e relevantes, otimizando o trabalho do foneticista,

descrevendo cada elemento com base no conhecimento de técnicas arquivísticas.

O padrão fonético descrito será representado em forma de matriz, pois é uma representação gráfica que permite a distribuição das informações e visibilidade da estrutura dos elementos. O resultado é entregue de volta ao foneticista, que é o responsável por lançar a matriz no sistema de identificação de falantes e realizar o processo de comparação (Figura 2). Cada matriz corresponde a um padrão fonético, que sob o ponto de vista da Arquivística pode ser considerado um dossiê, pois agrupa um conjunto de informações de um mesmo indivíduo.

Figura 2 - Esquema representativo das etapas do sistema de identificação de falantes



Fonte: Elaborado pelos autores

Os padrões fonéticos são considerados, conforme sua natureza, arquivos especiais, pois custodiam documentos de formas físicas próprias, que merecem tratamento especial no seu armazenamento, acondicionamento, registro, controle, conservação, entre outros procedimentos técnicos. Como se trata de fato relacionado à Justiça, pois o padrão fonético apresenta valor de prova para determinar a autoria de uma voz, o resultado desta aplicação, conforme sua espécie, pode ser considerada como atos de assentamentos. Atos de assentamentos são os configurados por registro consubstanciando assentamento sobre fatos ou ocorrências. (BELLOTTO, 2006, p.49)

Nessa perspectiva, a Arquivística pode fornecer os métodos necessários para descrever os padrões fonéticos, de modo a formar arquivos aptos a atender eficientemente a sistemas de identificação de falantes.

3 DESCRIÇÃO APLICADA NA FONÉTICA FORENSE

O arquivo tem a finalidade principal de servir à administração, tornando disponíveis as informações contidas no acervo, sendo importante para a tomada de decisões nas instituições. É imprescindível que os documentos estejam dispostos de forma a servir ao usuário com precisão e rapidez.

A evolução do conhecimento e o desenvolvimento de novas formas de comunicação e organização das informações colocam novos desafios à Arquivologia. Conforme afirma Paes (2004), esse constante crescimento qualitativo e quantitativo, também chamado de “explosão da informação”, provocou a evolução e o aperfeiçoamento das técnicas de registros e análise dos documentos. O conceito de documento não está vinculado apenas à documentação, mas a tudo que incorpora o valor probatório e significativo para o indivíduo. O documento é “[...] qualquer indicação concreta ou simbólica, conservada ou registrada com a finalidade de representar, reconstituir ou provar um fenômeno físico ou intelectual” (CUNHA; CAVALCANTE, 2008, p.132).

Assim, não há dúvidas que o conceito de documento abrange os registros de padrões fonéticos, pois o conceito evolui de forma técnica e cultural para atender aos desafios de um mundo em mudanças. Por isso, os conceitos, princípios e métodos arquivísticos também podem ser testados para tratar esses documentos.

Dentre esses métodos está a classificação, que é um procedimento técnico e intelectual fundamental para a organização correta do acervo, facilitando o acesso e a recuperação dos documentos. Conforme afirma Gonçalves (1998, p.12), o objetivo da classificação é, basicamente, dar visibilidade, permitindo de uma forma eficaz a organização de cada elemento.

Uma técnica que é imprescindível à classificação é a avaliação. Sem ela seria muito difícil o processo de classificar, pois é necessária para conhecer todo o trâmite, atribuir valores e distinguir o ciclo de vida dos documentos. De acordo com Souza e Araújo Junior (2013, p. 149): “A classificação de documentos de arquivo tem três objetivos: manter o vínculo arquivístico,

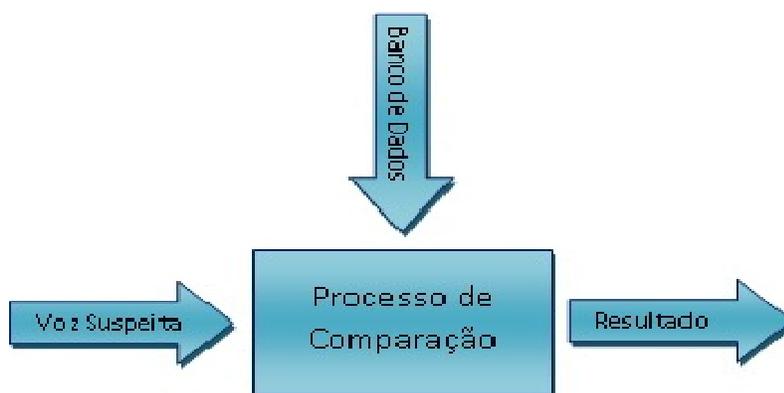
fundamentar a avaliação e possibilitar a recuperação da informação contida nos documentos.”

Sob uma perspectiva arquivística, a classificação aplica-se na gestão da informação em seus diversos suportes de forma a organizar as informações concretas ou abstratas para atender às necessidades dos indivíduos, incluindo-se, portanto, os padrões fonéticos, garantindo uma organização e recuperação da informação registrada.

Nesse sentido, sistemas de identificação de falantes podem ser desenvolvidos, conforme indicado na Figura 3, mediante a elaboração de modelos de descrição de padrões fonéticos criteriosamente organizados em bancos de dados, contendo, por exemplo, os padrões fonéticos extraídos de criminosos, que poderão ser identificados no caso do cometimento de crimes em que restem gravadas suas vozes.

Assim, para que o sistema de identificação de falantes seja utilizado de forma segura, é preciso que a descrição dos padrões fonéticos seja construída com todas as informações necessárias ao foneticista, organizando-se em matrizes. Desse modo, evita-se ouvir a gravação das falas ou reunir repetidamente dados fonéticos esparsos, possibilitando um acesso seguro e rápido, transformando os dados dos padrões fonéticos em uma linguagem documentária em forma de códigos e dados.

Figura 3 - Esquema representativo de sistema de identificação de falantes



Fonte: Elaborado pelos autores

A descrição, por sua vez, visa a uma recuperação rápida da informação através da atribuição de descritores (termos de indexação) a partir dos quais se possa encontrar o documento em buscas futuras.

Ao atribuir um descritor a um documento, o indexador declara que tal descritor possui alto grau de relevância para o conteúdo do documento; quer dizer, ele declara que o significado do descritor está fortemente associado a um conceito incorporado ao documento, e que é adequado à área temática do documento. (LANCASTER, 2004, p. 11).

Portanto, a descrição pode facilitar a recuperação da informação se aplicada de forma organizada e coerente. A representação das informações contidas nos padrões fonéticos possibilita uma expansão dessas informações quando disposta em matrizes que possam ser comparadas em sistemas de identificação de falantes e armazenadas em banco de dados. Externa-se, assim, a intencionalidade e razão da criação de documentos, ao representá-lo em uma linguagem documentária. Segundo Souza e Araújo Junior (2013, p. 156):

De modo geral, as linguagens naturais dificultam o tratamento da informação, causando problemas diversos relativos à imprecisão de termos, polissemia, considerações ambíguas sobre conceitos que impõem obstáculos à armazenagem e recuperação da informação. Logo, seu uso depende de uma tradução documentária, ou seja, a utilização de linguagens documentárias para mediar seu armazenamento e recuperação. Por isso, a escolha aleatória de termos não produz qualquer efeito de descrição dos conteúdos dos documentos.

As técnicas arquivísticas utilizadas para elaborar um modelo de matriz dos padrões fonéticos a serem comparados em sistema de identificação de falantes envolvem, portanto, a avaliação, a descrição e a indexação. E, para que essa gestão documental seja completa, é imprescindível a aplicação de uma tabela de temporalidade, para evitar um acúmulo de informações desnecessárias. Conforme Bernardes e Delatorre (2008, p. 36), tabela de temporalidade: “É o instrumento de gestão, resultante da avaliação documental, aprovado pela autoridade competente, que define prazos de guarda e a destinação de cada série documental, determinando sua preservação e sua eliminação.” O conceito pode ser aplicado aos padrões

fonéticos para determinar o ciclo documental conforme a idade do falante, devido às alterações que a voz apresenta durante a vida ou alterações fisiológicas do falante.

As alterações que ocorrem com a voz de todo falante, sobretudo por motivo da idade, impõem que se estabeleçam faixas etárias compatíveis para comparações de padrões fonéticos. Falantes com faixas etárias muito distintas têm suas vozes incompatíveis para confronto, devido às mudanças ocorridas nos seus aparelhos de fala. Assim, os padrões fonéticos devem obedecer a uma tabela de temporalidade baseada na adequação para uso comparativo no sistema de identificação, sendo classificados em correntes, intermediários e permanentes.

4 MODELO DE DESCRIÇÃO QUE REPRESENTA O PADRÃO FONÉTICO

Os padrões fonéticos são compostos de dados que precisam ser transformados pelo arquivista para tornarem-se aptos a serem comparados estatisticamente no sistema de identificação de falantes. Trata-se de elaborar um modelo adequado que agregue todas as informações dos padrões fonéticos de forma organizada, para serem armazenados no sistema e que sirvam para comparação estatística diretamente, sem a necessidade de outros processamentos. Nessa perspectiva, considerando que cada padrão fonético é composto de um conjunto de dados associados, que necessita ser tratado para ser processado num sistema informatizado, verifica-se que o formato adequado é o matricial.

O modelo matricial permite que os dados sejam eficientemente descritos e ainda facilmente processados em sistemas de análise estatística. Nesse sentido, escolheu-se como modelo específico a matriz-quadrada, com mesmo número de linhas e colunas, por apresentar propriedades especiais que facilitam as operações e o tratamento das informações no sistema. Ainda, fixou-se como tamanho a matriz com 10 linhas e 10 colunas (A10x10), por possuir elementos suficientes para a distribuição dos dados. No formato de

matriz, as informações precisam ser divididas de forma padronizada em dois campos: cabeçalho e dados (Figura 4)

Figura 4 - Modelo de matriz que representa os padrões fonéticos

COD 01	COD 02	COD 03	COD 04	COD 05	COD 06	COD 07	COD 08	COD 09	COD 10
COD 11	COD 22	COD 13	COD 14	COD 15	COD 16	COD 17	COD 18	COD 19	COD 20
a _{3,1}	a _{3,2}	a _{3,3}	a _{3,4}	a _{3,5}	a _{3,6}	a _{3,7}	a _{3,8}	a _{3,9}	a _{3,10}
a _{4,1}	a _{4,2}	a _{4,3}	a _{4,4}	a _{4,5}	a _{4,6}	a _{4,7}	a _{4,8}	a _{4,9}	a _{4,10}
a _{5,1}	a _{5,2}	a _{5,3}	a _{5,4}	a _{5,5}	a _{5,6}	a _{5,7}	a _{5,8}	a _{5,9}	a _{5,10}
a _{6,1}	a _{6,2}	a _{6,3}	a _{6,4}	a _{6,5}	a _{6,6}	a _{6,7}	a _{6,8}	a _{6,9}	a _{6,10}
a _{7,1}	a _{7,2}	a _{7,3}	a _{7,4}	a _{7,5}	a _{7,6}	a _{7,7}	a _{7,8}	a _{7,9}	a _{7,10}
a _{8,1}	a _{8,2}	a _{8,3}	a _{8,4}	a _{8,5}	a _{8,6}	a _{8,7}	a _{8,8}	a _{8,9}	a _{8,10}
a _{9,1}	a _{9,2}	a _{9,3}	a _{9,4}	a _{9,5}	a _{9,6}	a _{9,7}	a _{9,8}	a _{9,9}	a _{9,10}
a _{10,1}	a _{10,2}	a _{10,3}	a _{10,4}	a _{10,5}	a _{10,6}	a _{10,7}	a _{10,8}	a _{10,9}	a _{10,10}

Fonte: Elaborado pelos autores

4.1 Campo de cabeçalho da matriz

O campo de cabeçalho ocupa as duas primeiras linhas da matriz e é composto de informações que descrevem os dados contidos no documento, além de fornecer outras informações importantes para o processamento desses dados no sistema.

As informações do campo de cabeçalho devem ser preenchidas a partir de tabelas e outras informações elaboradas pelo arquivista, conforme especificado na Tabela 1.

Tabela 1 - Tabela de Códigos do Campo de Cabeço do Documento-Matriz

Código	Definição do Código	Referência
Código 01	Tipo de padrão fonético	Tabela 11
Código 02	Tipo de dado	Tabela 12
Código 03	Tipo de amostra	Tabela 13
Código 04	Data da coleta	Tabela 14
Código 05	Sexo do falante	Tabela 15
Código 06	Número de registro do falante no sistema	–
Código 07	Número de registro da amostra no sistema	–
Código 08	Ano de nascimento do falante	–
Código 09	Faixa etária do falante na data da coleta	Tabela 16
Código 10	Período do documento em arquivo corrente	
Código 11	Período do documento em arquivo intermediário	
Código 12	Período do documento em arquivo permanente	
Códigos de 13 a 20	Códigos reservados para novas informações	Índice Q

Fonte: Elaborado pelos autores

O Código 01 da Tabela de Códigos do Campo de Cabeçalho indica o tipo do padrão fonético que está representado na matriz, cujos índices estão definidos na Tabela 2.

Tabela 2 - COD 01

Índice	Definição do Índice
1	Padrão fonético acústico: Dispersão de Vogais
2	Padrão fonético acústico: Dispersão de Ditongos
3	Padrão fonético acústico: Distribuição de f_0
4	Padrão fonético acústico: Distribuição Espectral
5	Padrão fonético acústico: Distribuição de Formantes
6	Padrão fonético Perceptual

Fonte: elaborado pelos autores

O Código 02 da Tabela de Códigos do Campo de Cabeçalho indica o tipo de dado do padrão fonético que está representado na matriz, cujos índices estão definidos na Tabela 3.

Tabela 3 - COD 02

Índice	Definição do Índice
1	Contínuo normal
2	Contínuo Anormal
3	Categórico

Fonte: elaborado pelos autores

O Código 03 da Tabela de Códigos do Campo de Cabeçalho indica o tipo de amostra do padrão fonético que está representado na matriz, cujos índices estão definidos na Tabela 4.

Tabela 4 - COD 03

Índice	Definição do Índice
1	Dependente
2	Independente

Fonte: elaborado pelos autores

O Código 04 da Tabela de Códigos do Campo de Cabeçalho indica a data da coleta do padrão fonético, com formato “ddmmaaaa”, especificado na Tabela 5. A data da coleta do padrão fonético tem especial importância, pois servirá de referência para definir a situação do documento quanto à temporalidade, como será discutida mais adiante.

Tabela 5 - COD 04 – indica a data da coleta do padrão fonético representado na matriz

Formato da Data	dd	mm	aaaa
ddmmaaaa	dia	mês	ano

Fonte: elaborado pelos autores

O Código 05 da Tabela de Códigos do Campo de Cabeçalho indica o sexo do falante, cujos índices estão definidos na Tabela 6.

Tabela 6 - COD 05 – indica o sexo do falante

Índice	Definição do Índice
1	Masculino
2	Feminino

Fonte: elaborado pelos autores

O Código 06 da Tabela de Códigos do Campo de Cabeçalho indica o número do registro em que o falante está cadastrado no sistema, permitindo, assim, obterem-se suas informações detalhadas, como nome, endereço, telefone etc.

O Código 07 da Tabela de Códigos do Campo de Cabeçalho indica o número do registro da amostra do padrão fonético cadastrada no sistema, para efeito de indexação e endereçamento do processamento das informações

O Código 08 da Tabela de Códigos do Campo de Cabeçalho indica o ano de nascimento do falante, que servirá para estabelecer a situação do documento quanto à temporalidade.

Código 09 da Tabela de Códigos do Campo de Cabeçalho indica a faixa etária do falante no momento da coleta do padrão fonético, especificado na Tabela 7.

Tabela 7 - COD 09 – Faixa etária do falante em anos

Índice	Faixa	Índice	Faixa	Índice	Faixa	Índice	Faixa
1	0 – 0,5	6	8 – 12	11	33 – 40	16	85 – 90
2	0,5 – 1	7	12 – 13	12	40 – 60	17	90 – 95
3	1 – 3	8	13 – 16	13	60 – 70	18	95 – 100
4	3 – 5	9	16 – 21	14	70 – 80	19	100 – 105
5	5 - 8	10	21 – 33	15	80 – 85	20	105 – 110
5	5 - 8	10	21 – 33	15	80 – 85	20	105 – 110

Fonte: elaborado pelos autores

A tabela 7 indica as faixas etárias e servirá de referência para estabelecer a situação do documento quanto à temporalidade.

4.2 Campo de dados da matriz

O campo de dados ocupa desde a terceira até a décima linha da matriz e é formado pelos dados do padrão fonético, distribuídos como elementos da matriz, abaixo do cabeçalho. A figura 5 apresenta o modelo de campo da matriz.

Figura 5 - Modelo de campo da matriz

a _{3,1}	a _{3,2}	a _{3,3}	a _{3,4}	a _{3,5}	a _{3,6}	a _{3,7}	a _{3,8}	a _{3,9}	a _{3,10}
a _{4,1}	a _{4,2}	a _{4,3}	a _{4,4}	a _{4,5}	a _{4,6}	a _{4,7}	a _{4,8}	a _{4,9}	a _{4,10}
a _{5,1}	a _{5,2}	a _{5,3}	a _{5,4}	a _{5,5}	a _{5,6}	a _{5,7}	a _{5,8}	a _{5,9}	a _{5,10}
a _{6,1}	a _{6,2}	a _{6,3}	a _{6,4}	a _{6,5}	a _{6,6}	a _{6,7}	a _{6,8}	a _{6,9}	a _{6,10}
a _{7,1}	a _{7,2}	a _{7,3}	a _{7,4}	a _{7,5}	a _{7,6}	a _{7,7}	a _{7,8}	a _{7,9}	a _{7,10}
a _{8,1}	a _{8,2}	a _{8,3}	a _{8,4}	a _{8,5}	a _{8,6}	a _{8,7}	a _{8,8}	a _{8,9}	a _{8,10}
a _{9,1}	a _{9,2}	a _{9,3}	a _{9,4}	a _{9,5}	a _{9,6}	a _{9,7}	a _{9,8}	a _{9,9}	a _{9,10}
a _{10,1}	a _{10,2}	a _{10,3}	a _{10,4}	a _{10,5}	a _{10,6}	a _{10,7}	a _{10,8}	a _{10,9}	a _{10,10}

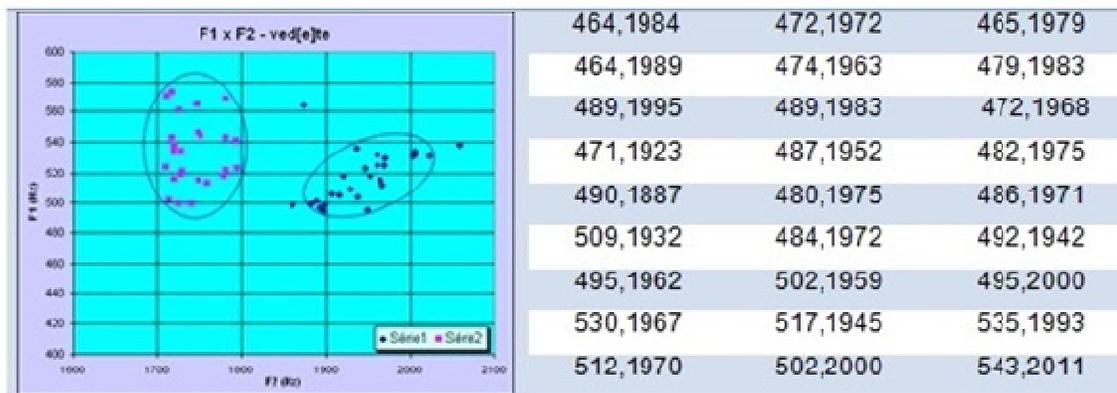
Fonte: elaborado pelos autores

Para exemplificar, será elaborado a descrição do documento que representa o padrão acústico de dispersão de vogais, com pares ordenados de formantes (F1, F2), mostrado na figura 6.

Para a escolha da análise estatística adequada, os dados de cada padrão fonético precisam ser classificados quanto:

1. ao tipo de dado: contínuo ou categórico;
2. à distribuição dos dados: normal ou anormal e;
3. ao tipo de amostra: dependente ou independente.

Figura 6 - Padrões acústicos de dispersão de vogais (F1, F2) Hz



Fonte: Braid, 2010, p.10

Os padrões fonéticos acústicos são: dispersão de vogais, dispersão de ditongos, distribuição de fo, distribuição espectral e distribuição de formantes. Todos esses padrões acústicos correspondem a parâmetros numéricos que podem ser medidos em uma escala contínua, extraídos de valores de frequência e energia da voz do falante. Portanto, os dados desses padrões são

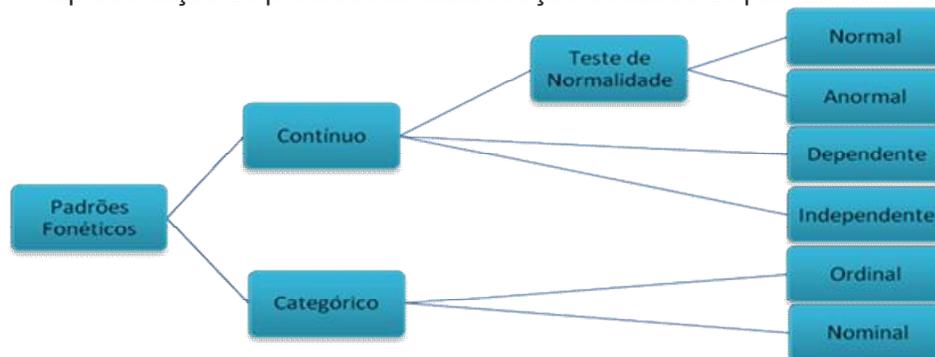
números que podem ser fracionados e, por isso, são descritos como contínuos ou quantitativos. Os padrões fonéticos acústicos podem apresentar dois tipos básicos de distribuição dos dados: normal e anormal.

Dependendo da maneira como os padrões fonéticos são fornecidos ao arquivista, classificar os dados quanto à distribuição pode ser uma tarefa complexa, no entanto testes estatísticos podem ser utilizados para a verificação da normalidade.

Além da descrição dos dados dos padrões fonéticos quanto ao tipo e distribuição, para a adequação do tratamento estatístico também é necessário determinar se as amostras coletadas são dependentes ou independentes.

Assim, o processo de classificação dos dados dos padrões fonéticos, com o objetivo de prepará-los para análises estatísticas pode seguir a sequência apresentada no esquema da Figura 7.

Figura 7 - Representação do processo de classificação de dados de padrões fonéticos

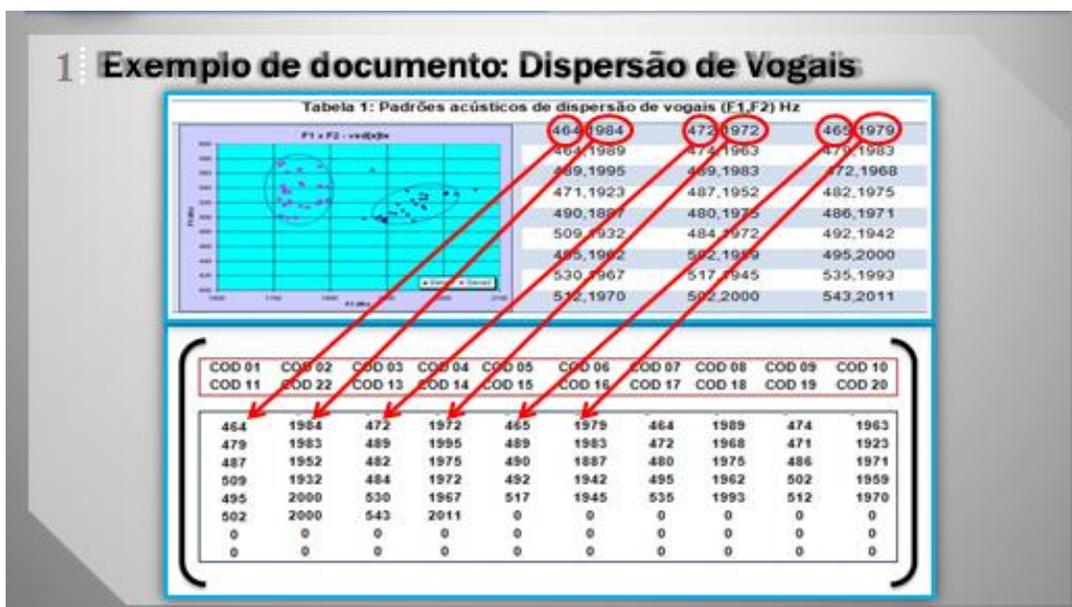


Fonte: adaptação dos autores a partir dos dados obtidos verbalmente com César Braid

4.3 Modelo do Padrão Fonético transformado em matriz

A Figura 6 apresentou um exemplo de padrão fonético de dispersão de vogais que passará pela classificação descrita acima. A Figura 8 apresenta as informações do padrão fonético transferido para a matriz.

Figura 8 - Modelo de matriz que representa os padrões fonéticos



Fonte: elaborado pelos autores

A Figura 9 apresenta um exemplo da representação do cabeçalho do padrão fonético de dispersão de vogais, que foi descrito na seção 4.1.

Figura 9 – Exemplo de cabeçalho

COD 01	COD 02	COD 03	COD 04	COD 05	COD 06	COD 07	COD 08	COD 09	COD 10
COD 11	COD 22	COD 13	COD 14	COD 15	COD 16	COD 17	COD 18	COD 19	COD 20

1	1	2	12042015	1	002548	032158	1991	10	2024
20252051	2052	0	0	0	0	0	0	0	0

Fonte: elaborado pelos autores

A Figura 10 apresenta um exemplo de matriz completa com os dados fonéticos transferidos após a classificação organizada pelo arquivista. A matriz representa o padrão acústico e contém todas as informações necessárias para o processamento em sistemas de identificação de falantes.

Figura 10 - Exemplo de descrição que representa o padrão acústico de dispersão de vogais

1	1	2	12042015	1	002548	032158	1991	10	2024
20252051	2052	0	0	0	0	0	0	0	0
464	1984	472	1972	465	1979	464	1989	474	1963
479	1983	489	1995	489	1983	472	1968	471	1923
487	1952	482	1975	490	1887	480	1975	486	1971
509	1932	484	1972	492	1942	495	1962	502	1959
495	2000	530	1967	517	1945	535	1993	512	1970
502	2000	543	2011	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Fonte: Elaborado pelos autores

4.4 Tabela de Temporalidade dos Padrões Fonéticos

A voz humana sofre alterações ao longo da idade do falante, devido a mudanças fisiológicas que ocorrem no aparelho de fala, modificando os padrões fonéticos. Assim, existe plena adequação na comparação de duas vozes que apresentam as mesmas faixas etárias. Todavia, o confronto indiscriminado de padrões fonéticos de falantes em faixas etárias distintas tende a dar resultados não confiáveis. Nessa perspectiva, é importante que se estabeleça uma tabela de temporalidade tomando como base os índices de faixas etárias da Tabela 7, para que sirva de referência na análise de compatibilidade comparativa entre documentos.

A tabela de temporalidade servirá para determinar quais documentos estão prontamente adequados para a comparação, portanto retidos em arquivos correntes, e aqueles que discriminada e circunstancialmente poderão ser comparados em uma segunda fase do processamento por estarem em arquivos intermediários. O restante dos documentos, todos incompatíveis para o exame comparativo, deverão ser destinados a arquivos permanentes, pois são elementos forenses que podem ter sido usados em outras comparações e, portanto, não devem ser descartados.

Tendo-se como referência os índices de faixa etária dispostos na Tabela 7, a temporalidade dos documentos obedece aos critérios mostrados na Tabela 8, sendo $I_1 \leq I_2$ e definidos como:

- índices de falantes 1 e 2 que terão os padrões fonéticos comparados;

- Índices do mesmo falante que mudou de faixa etária, de I_1 para I_2 . Trata-se da situação em que o documento foi criado quando o falante tinha faixa etária do índice I_1 e, após alguns anos de arquivamento, passa para o índice I_2 . Nesse caso, I_2 é determinado tomando-se como referência a idade atualizada do falante.

Tabela 8 - Temporalidade dos Documentos

	Arquivo Corrente	Arquivo Intermediário	Arquivo Permanente
Condição	$I_2 - I_1 = 0$	$I_2 - I_1 \leq 2$	$I_2 - I_1 > 2$

Fonte: Elaborado pelos autores

6 CONCLUSÃO

A identificação de falantes é a área da Fonética Forense capaz de determinar a autoria de uma voz por meio da comparação de padrões fonéticos de indivíduos. É nessa perspectiva que algoritmos de sistemas de identificação de falantes são desenvolvidos, todavia é necessário que os padrões fonéticos sejam descritos em formato adequado para serem comparados e recuperados. Os padrões fonéticos diretamente extraídos pelo foneticista correspondem a valores quantitativos e qualitativos, mas ainda estão impróprios para as análises nos algoritmos, porque precisam de tratamento adequado que os represente em linguagem documentária.

A Arquivologia detém o conhecimento necessário para representar a informação em linguagem documentária, podendo, portanto, ser aplicada em diversas atividades que precisam da gestão de informação, como a Fonética Forense. As técnicas da Arquivística para avaliar, classificar, organizar, descrever e indexar a informação possibilitam a adequada armazenagem e recuperação desses documentos.

Os padrões fonéticos coletados de um indivíduo correspondem a registros documentais e, por isso, a aplicação dos conhecimentos arquivísticos no processamento documental desses fenômenos da fala, para a utilização em sistemas de identificação de falantes, traduz-se na profissionalização dos

meios para adequar essas informações visando à otimização de sua armazenagem, recuperação e análise de temporalidade.

Este trabalho demonstrou como a Arquivologia pode auxiliar a Fonética Forense com a descrição de padrões fonéticos para o desenvolvimento de sistemas de identificação de falantes. Para compreender o papel da Arquivologia nesses sistemas é necessário ressaltar que o conhecimento arquivístico, evidentemente, não pretende extrair os padrões fonéticos da fala de indivíduos ou ainda desenvolver os métodos de processamento da tecnologia da informação utilizados na identificação de falantes, mas elaborar a descrição dos padrões com informações fornecidas pelo foneticista.

A pesquisa desenvolveu-se com o objetivo de elaborar o modelo de descrição adequada para os padrões fonéticos, agregando todas as informações organizadamente para serem armazenados e recuperados no sistema visando à realização de comparações. Como o padrão fonético é um conjunto de dados associados, que deverá ser processado num sistema informatizado, concluiu-se que o seu formato representativo adequado é o matricial.

A aplicação das técnicas da Arquivologia, por meio da avaliação, descrição e indexação, permitiu tratar as informações dos padrões fonéticos, de modo a cooperar significativamente com o trabalho do foneticista, transformando os dados em linguagem documentária acessível e facilitando o processo de comparação em sistemas de identificação de falantes.

Observa-se que a aplicação do conhecimento arquivístico, conforme desenvolvido neste trabalho, pode ser usado no tratamento de outras informações com aplicação em diversos tipos de sistemas e áreas. Essa interdisciplinaridade é muito importante para ampliar a aplicabilidade da Arquivística no mercado profissional, repercutindo no reconhecimento, na visibilidade e na gestão da informação.

O estudo também suscita futuras pesquisas que podem favorecer ainda mais as duas áreas, sobretudo no desenvolvimento de sistemas completos de identificação de falantes, de forma a manter a agilidade, o acesso à informação, a produtividade e o menor custo de operação dos sistemas.

REFERÊNCIAS

BELLOTTO, Heloísa Liberalli. **Arquivos Permanentes**: tratamento documental. 4.ed. Rio de Janeiro: FGV, 2006. .

BERNADES, Ieda Pimenta; DELATORRE, Hilda. **Gestão Documental Aplicada**. São Paulo: Arquivo Público do Estado de São Paulo, 2008.

BRAID, Antonio César Morant **Fonética Forense**. 2. ed. Campinas: Millennium, 2003.

_____. Fundamental frequency alterations due to read speech and increased vocal effort. **Australian and New Zealand Forensic Sciences Society**, 2010.

GONÇALVES, Janice. **Como classificar e ordenar documentos de arquivo**. São Paulo: Arquivo do Estado, 1998.

LANCASTER, F.W. **Indexação e Resumos**: teoria e prática. Brasília: Brinquet de Lemos, 2004.

PAES, Marilena Leite. **Arquivo**: teoria e prática. 3. ed. Rio de Janeiro: FGV, 2004.

Title

The archival knowledge applied in the description of phonetic patterns of speakers

Abstract:

Introduction: Speaker identification is the activity that is able to determine if the same speaker produced two voices. The phonetician performs the exam by comparing the extracted phonetic patterns of speakers, which are elements that determine the speech traits of an individual. **Objective:** This study shows how the archival knowledge can be used to organize the phonetic patterns provided by phonetician, helping to build speaker identification systems. **Methodology:** The research is characterized as descriptive and qualitative and quantitative, with the interpretation, description, evaluation and indexing of phonetic patterns, besides the application of statistical concepts. **Findings:** The results of the investigation led to the proposal of a model description of the phonetic patterns, organized in a matrix form, with relevant information to the phonetician to be used in speaker identification systems. **Conclusion:** It is concluded that the practice of interdisciplinary of the two areas, Archivology and Phonetics, can make an important contribution in the results of organizing data and phonetic information, demonstrating the broad scope of archival knowledge on current problem-solving.

Keywords: Archival description. Information management. Forensic Phonetics.

Titulo

El conocimiento archivístico aplicado en la descripción de estándares fonéticos

Introducción: La identificación del hablante es el proceso para determinar si dos voces fueron emitidas por la misma persona. La prueba es realizada por el fonetista, mediante la comparación de los patrones fonéticos extraídos de los hablantes, que son los elementos que determinan rasgos característicos de la voz de un individuo.

Objetivo: Este estudio pretende demostrar cómo el conocimiento de archivos se puede utilizar para describir los patrones fonéticos proporcionadas por el fonetista, ayudando a construir sistemas de identificación de hablantes.

Metodología: La investigación se caracteriza por ser descriptiva, cualitativa y cuantitativa, con la interpretación, descripción, evaluación e indización de los patrones fonéticos, además de la aplicación de los conceptos estadísticos.

Resultado: El resultado de la investigación condujo a la propuesta de un modelo de descripción de los patrones fonéticos, organizadas en forma de matriz, con la información pertinente a la fonética para ser utilizado en sistemas de identificación de los hablantes.

Conclusión: Se concluye que la práctica de la interdisciplinariedad entre las dos áreas, Archivología y Fonética, puede contribuir de manera importante en los resultados de la organización de los datos y la naturaleza de la información auditiva, lo que demuestra la amplia gama de conocimientos de archivos para resolver los problemas de hoy en día.

Palabras clave: Descripción de archivos. Gestión de la información. Fonética Forense.

Recebido: 12.02.2016

Aceito: 25.03.2017