

A UNIDADE MÍNIMA DE PERCEPÇÃO DA FALA É A SILABA OU O FONEMA?

REGINA JAKUBOVICZ (*)

(*) Fonoaudióloga, Doutora em Fonoaudiologia pela Universidade do Museu Social Argentino – Buenos Aires – Argentina; Especialista em Linguagem pelo CFFa (Conselho Federal de Fonoaudiologia); Professora Titular no curso de Graduação da Universidade Estácio de Sá – Rio de Janeiro – Brasil; Professora no curso de Pós-graduação da Universidade Veiga de Almeida – Rio de Janeiro – Brasil (Regina.jakubovicz@infolink.com.br).

RESUMO

Neste artigo é feita uma revisão das teorias de percepção mínima da fala desde suas origens. Vários estudos sobre a percepção acústica são expostos. Algumas hipóteses são colocadas e analisadas. Na discussão, faz-se uma revisão e resumo dos conceitos vistos. Na conclusão mostra-se a importância e a necessidade de que tal assunto seja estudado e pesquisado já que ele seria importante para um maior conhecimento de vários distúrbios da fala oral e escrita e seu respectivo tratamento.

PALAVRAS – CHAVE: Percepção; sílaba; fonema; Terapia da Fala.

ABSTRACT

The author presents a review of the theories of minimal units for speech perception since their origins. Several studies on acoustic perception are exposed. Some hypotheses are put and analyzed. In the discussion, the main concepts are reviewed and summarized. In the conclusion, one shows the importance and necessity of such matter to be studied and researched, because this would be important for a better understanding of various disorders of the oral and written speech and their treatment.

KEY-WORDS: Perception; syllable, phoneme, Speech Therapy.

1. INTRODUÇÃO

Não resta dúvida que o sistema auditivo é sintonizado para os sons da fala. Tudo indica então que na evolução os mecanismos da fala e o sistema auditivo desenvolveram-se em paralelo, mas de uma maneira tão unida que somos seres afiados em ouvir os sons que nós mesmos produzimos. Apesar de sermos especialmente dotados para perceber os sons que emitimos, a tarefa de desvendar os processos envolvidos na percepção da fala é das mais difíceis ou quase impossível. Qual seria a unidade mínima de percepção da fala; é ao nível da sílaba ou do fonema? Os sons da fala raramente são produzidos de maneira isolada, eles se influenciam um ao outro dando como resultado uma produção oral. Quando Jakobson (1956), elaborou o modelo binário em que cada unidade fonêmica contrastava e se opunha a outra, ele colocou o fonema como unidade de base da percepção. Chomsky (1968) modificou este modelo acrescentando aspectos articulatórios muito mais do que acústicos. O sistema de Chomsky e Halle (1968) é baseado primeiramente na estrutura da língua e só secundariamente na fisiologia da fala.

Existem tres principais teorias sobre a percepção da fala:

- 1 - "motor theory" (teoria motora) de Liberman (1957)
- 2 - "analysis by synthesis" (análise pela síntese) de Halle (1956)
- 3 - "passive theory" (teoria passiva) de Fant G. (1967)

A teoria motora postula que o mecanismo usado pelos ouvintes para assimilar o sinal acústico do fonema é o conhecimento que ele tem das regras dos movimentos articulatórios. O ouvinte não percebe combinações de sons que ele não conhece ou que não pode realizar. Outro aspecto que a teoria motora mostra é que a fala tem propriedades perceptuais especiais o que a faz ser percebida de maneira diferente de outros sons.

Na teoria de análise pela síntese o ouvinte também se refere à sua produção, mas essa referencia é mais acústica, menos articulatória e se baseia num sistema de combinação. Isto quer dizer que o ouvinte ouvirá a mensagem e irá analisá-la evocando o modelo auditivo de sua própria produção. Haverá síntese neural ou de neurônios se o modelo ouvido casar ou combinar com o padrão cerebral. Nesse caso há a aceitação do modelo ou a percepção.

A teoria passiva de Fant coloca a percepção da fala primordialmente como sensorial. Segundo ele, os mecanismos perceptuais compartilham com os mecanismos de produção características distintas mas em comum. A diferença fundamental é que o ouvinte não necessita recorrer à produção para perceber os sons. Os centros linguísticos no cérebro são comuns tanto para receber como para produzir mensagens, mas os centros responsaveis pelos subfonemas, nos seus aspectos perifericos, podem ser vistos como independentes.

A teoria motora tem a falha de ser difícil encontrar na atividade dos músculos uma correspondência direta entre o fonema e o sinal acústico. Na teoria de análise e síntese existe a vantagem de que o ouvinte utiliza os seus conhecimentos das regras fonológicas, sendo capaz então de analisar as variações feitas em alta velocidade ou com distorções.

A teoria passiva parte do princípio que o ouvinte pode decodificar diretamente os sons se eles fizerem parte de sua produção (a língua natal), caso contrário a tarefa de percepção torna-se muito mais difícil, o que não deixa de ser verdade.

Como é realizada a fala? Possivelmente ela se origina através de impulsos neurais. Estes impulsos produzem tensão nos músculos da face, língua, laringe e no sistema respiratório. A tensão no músculo ocasiona uma trajetória nos órgãos articulatórios de acordo com as leis da mecânica. O som, já modificado, é transmitido através de ondas sonoras ao ouvido. Uma vez transmitido, ele é convertido em pontos específicos ao longo da membrana basilar. Estes movimentos, que estão diretamente relacionados à frequência dos componentes da onda sonora, resultam finalmente no impulso nervoso que acontece no VIII par craniano. Logo numa articulação existe um deslocamento em sequência dos articuladores, a passagem da onda sonora e um impulso neural no ouvido. Em essência, um ato articulatório consiste numa sequência de posições com alvos a atingir.

Existe pouca variação entre o sinal acústico e o fonema, como entidade em si. Liberman (1970) diz que: "a informação acústica correspondente a um fonema muitas vezes está distribuída em vários segmentos. Na fala normal podem ser produzidos de 10 a 15 fonemas num segundo e muitos fonemas tem a duração de 50 milissegundos (msec.) ou menos (Liberman et al. 1967). Quando ouvimos um som produzido por nós mesmos, temos a impressão que a recepção e a produção oro-sensorial dos articuladores estão em paralelo, mas isto não quer dizer que temos percepção imediata. O controle da produção da fala não é feito no mesmo momento da percepção, como seria lógico supor. Para que houvesse um controle imediato, o mecanismo sensorial ou de recepção da fala teria de ser capaz de captar, analisar e corrigir de zero a zero, ou seja, teria de ter uma capacidade temporal ultraveloz. Existe então o tempo de resposta reflexa (resposta neural) que é o tempo que leva o músculo interno intercostal para entrar em atividade assim que começamos a falar. Temos também o tempo de resposta voluntária (resposta mecânica) que é o aumento da pressão subglótica responsável pelo início da atividade de produção nas cordas vocais. O tempo de resposta neural ou resposta do músculo interno intercostal é estimado de ser em torno de 140 a 320 msec. e a resposta mecânica, que é o início da fonação, é estimado de ser de 48 a 60 msec. Logo nós não ouvimos ao mesmo tempo em que articulamos a resposta, o que nos indica que o controle oro-sensorial ou a retroalimentação é muito mais feito no sentido de determinar se a produção da fala foi feita, do que se ela está sendo feita. McNeill et al. (1973) descobriram que se leva cerca de 70 msec. a mais para identificar uma oclusiva do que para produzi-la e muito mais tempo ainda para identificar uma articulação errada, reorganizar e corrigir. Massaro em 1974 mostrou a necessidade de um período de 250 msec para processar e armazenar a unidade correspondente a uma sílaba. Tais

fatos talvez nos levam a caminhar mais na direção de aceitarmos a sílaba como a base da percepção da fala muito mais do que o fonema. Além do mais, a sílaba é reconhecivelmente muito mais necessária na identificação da palavra.

Existem poucas informações de como os indivíduos percebem os sons da fala e isso se deve ao fato da percepção da palavra ter como referentes a resposta dos articuladores, como dizem algumas teorias e não o sistema auditivo como um todo. Jerger (1963), no seu livro *Modern Developments in Audiology*, já dizia: "O que mais dificulta desvendar o problema da percepção auditiva é considerar-se que:

- Sons puros são estímulos suficientes para se testar a recepção do complexo sistema auditivo.
- As vias que integram a audição serem só passivas, isto é, apenas condutoras de som.
- Ignorar-se o papel exercido pelo sistema auditivo como um todo, ou seja, a maneira como é processada, armazenada e transmitida a informação auditiva através do sistema.

Como vemos, já naquela época, Jerger nos falava da importância dos estudos em processamento auditivo central, o que se tornou uma realidade bem recente.

2. PESQUISAS SOBRE A PERCEPÇÃO DA FALA

Em 1974 Liberman e colaboradores, avaliaram crianças de 4 a 7 anos utilizando-se de uma espécie de jogo em que as elas deveriam identificar o número de toques silábicos e de sons vocálicos. Só crianças de 6 e 7 anos puderam realizar bem o reconhecimento dos sons e das sílabas. Crianças de 4 e 5 anos só reconheciam a sílaba. Baseados nesses resultados, Morais J. et al. (1979), levantaram dúvidas sobre os resultados alcançados, achando que eles aconteceram devido ao amadurecimento cognitivo ou ao treino envolvido no processo de aprender a ler e a escrever a que são submetidas crianças de 6 a 7 anos. Morais e seus colaboradores mostraram também que adultos analfabetos obtinham os mesmos resultados das crianças de 4 e 5 anos na identificação de toques silábicos e sons vocálicos; ou seja, elas reconheciam apenas a sílaba. A conclusão de Morais é que a sílaba é a unidade disponível para uma classificação auditiva simples, enquanto que o fonema está mais disponível na percepção daqueles que dominam o processo correspondente entre o fonema e o grafema.

O papel da sílaba na percepção da fala foi explorado também nos estudos de testes perceptuais por Savin e Bever (1970). Eles fizeram uma pesquisa em que pediam a sujeitos para detectar, o mais rápido possível, o primeiro fonema ou a sílaba apresentada numa lista de palavras sem sentido. Os resultados obtidos levaram os autores a concluir que os fonemas não são percebidos diretamente, mas são derivados da análise da percepção da síntese silábica.

Ambos os estudos citados acima sofreram críticas, sendo que o que mais se questionou foi a distinção entre percepção e identificação. Foss e Swinney (1973) mantiveram que: “geralmente, pequenas unidades são facilmente identificadas quando se fraciona as grandes unidades”. Com esta hipótese eles afirmam que a detecção do fonema e da sílaba implica na identificação da palavra na qual o fonema ou a sílaba se encontra inserido. Tal fato, sujeito a comprovação empírica, foi analisado pelo próprio Foss e Blank em 1980. Os resultados obtidos tomaram caminhos diferentes dos estudos anteriores e podem ser resumidos da seguinte maneira: “a percepção ou a identificação de uma unidade linguística num nível (x) não é necessariamente derivado de (x + 1)”. Isto quer dizer que a percepção de um som vocálico isolado como a vogal /a/ não é derivado da percepção de uma síntese silábica como a junção d+a; /da/. Isto sugere que a percepção do som /da/ não é necessariamente o resultado da análise da palavra onde a sílaba está inserida como em /soldado/.

Para esclarecer melhor o papel da sílaba na segmentação da fala, Mehler, Dommergues e Frauenfelder (1981) escolheram seqüências fonêmicas que tinham ou não tinham correspondência na estrutura silábica das palavras estímulos a serem comparadas. Tomando um exemplo de seus estudos verificamos que: ambas as palavras PALACE e PALMIER, começam com o mesmo fonema /p a/ mas a localização dos fonemas /p a l/ e /p a/ nas palavras, fornecem uma pista acústica diferente. No experimento de Mehler e seus colaboradores, os sujeitos detectam visualmente com mais velocidade a seqüência de fonemas quando eles aparecem na primeira sílaba /p a/ do que quando aparecem na sílaba /p a l/. Uma análise do que os sujeitos fazem neste tipo de experiência forneceram pistas interessantes. Vejamos:

- ao verem o estímulo escrito as pessoas armazenam uma representação mental dele,
- subsequentemente ao ouvirem o sinal acústico eles fazem uma análise com a finalidade de reproduzir o estímulo,
- o processo de detecção só começa realmente a acontecer a partir do pareamento da representação mental com a análise acústica feita.

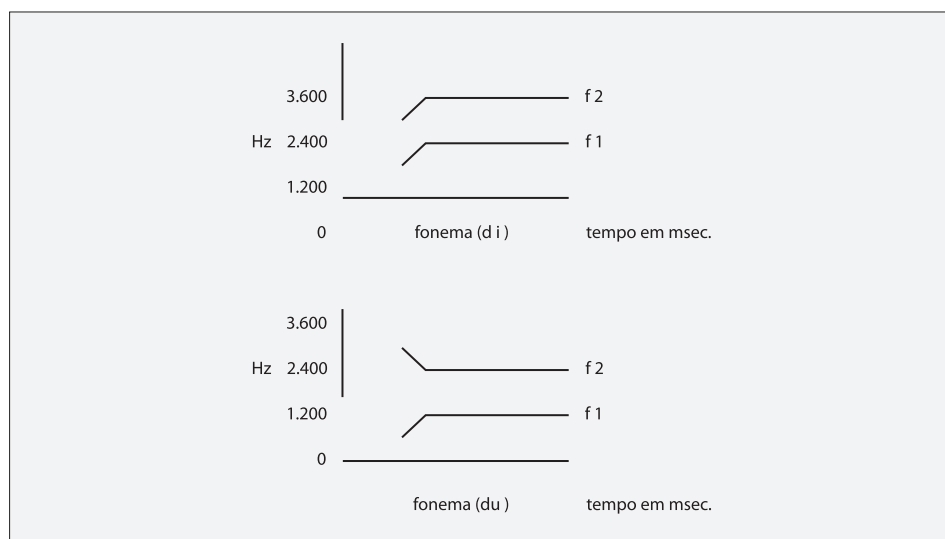
Essas hipóteses vêm ao encontro a uma proposta de Mills (1980) que dizia o seguinte: “quanto mais próxima for a expectativa do sujeito em relação ao alvo a distinguir, mais rápido ele responderá e fará o seu reconhecimento”. Mills pressupõe que exista no cérebro um padrão fisiológico para cada fonema, o que explica a expectativa dos sujeitos sobre o fonema ou a sílaba que virá a seguir e o menor tempo na identificação. A hipótese do padrão neural talvez tenha um fundo de verdade. Podemos verificar isso ao estudarmos e emissão de parafasias fonêmicas feitas por pessoas com afasia depois de um acidente vascular cerebral (AVC). As substituições fonêmicas do tipo; /otelha/ por /orelha/ ou /samonete/ por /sabonete/, acontecem sempre em fonemas muito próximos da palavra alvo, sugerindo que exista no cérebro, um padrão neural já impresso para cada fonema. O paciente com AVC não troca “orelha” por “elefante”, o que

quer dizer que a substituição se dá ao nível do fonema e não do significado da palavra. Como o cérebro após o AVC fica com áreas especializadas isoladas e com neurônios desconectados, é possível que não haja mais a transmissão correta da informação aos mecanismos motores encarregados de emitir a sequência fonêmica.

Foi a partir de estudos nos espectrogramas, que se tornou possível chegar mais perto da realidade acústica existente entre o fonema e a sílaba. Usando os espectrógrafos foi possível visualizar que são os formantes das vogais que determinam a transição acústica e não a consoante. Na análise espectrográfica verifica-se que não há correspondência ideal entre o fonema e o sinal acústico. Sabemos que os formantes são picos de frequência reforçados, o que quer dizer que eles possuem a propriedade acústica mais importante na análise do som. Nos espectrogramas vemos nitidamente que é a passagem da vogal à consoante ou a transição, que determina em que frequência (em Hertz) concentra-se a maior ou menor energia acústica (ver anexo).

Delattre (1955) estudou a transição consoante/vogal que ele chamou de LOCUS ou o local. Analisando as sílabas ele descobriu que não há correspondência entre o fonema e o sinal acústico. De acordo com a transição o mesmo fonema tem frequência diferente e dependendo do formante da vogal seguinte a sílaba será percebida de outra maneira.

Nos quadros abaixo temos o *Locus* das sílabas (di) e (du), onde se vê a mudança de frequência de 2.400 Hz para 3.600 Hz no formante 2 (F2) na sílaba (di) e a modificação para uma frequência mais baixa (3.600 para 2.400.Hz) em (F2) na sílaba (du). Os índices acústicos são pois condicionados pelo contexto onde o som está inserido, se isolado ou se na sílaba.



3. DISCUSSÃO

A revisão dos estudos sobre o tema ressalta os seguintes pontos teóricos:

- O reconhecimento da sílaba parece ter alguma relação com o conhecimento que o sujeito tem sobre o grafema e o fonema. Pessoas alfabetizadas armazenam a informação auditivamente e também visualmente. Sujeitos analfabetos aprendem os sons apenas de maneira auditiva. Para estes últimos, os fonemas possuem uma seqüência auditiva, não havendo uma representação mental deles. A percepção das crianças (não alfabetizadas) se dá da mesma maneira que os adultos analfabetos, isto é, só de maneira auditiva e sem representação mental.

- Os fonemas não são percebidos diretamente, o seu reconhecimento é a consequência da síntese silábica. Os fonemas não se seguem uns aos outros como pérolas num cordão, mas são produzidos e percebidos como eventos paralelos. No signo / du /, tanto o / d / como o / u / é influenciado um pelo outro e formam um contexto silábico que tem significação. A palavra percebida e ouvida por alguém, está em mediação de alguma maneira com o sistema de fala daquele que escuta o som. Escutar alemão é o mesmo que escutar ruídos com entonação para aquele que não fala alemão. Como os sons não estão no repertório de quem escuta, a pessoa não pode perceber e por isso mesmo não consegue emití-los. Parece então que o processo de recepção da fala seria feito por análise e síntese: a pessoa escutaria o som, analisaria, sintetizaria e responderia a partir daí. Nesta visão, a unidade de percepção da fala não seria o fonema, mas sim a síntese silábica.

- No livro "Vocabulário dos distúrbios da comunicação" de Nicolosi, Harryman e Kreshech (1989), o termo percepção auditiva é definido como: "identificação, interpretação ou organização de dados sensoriais recebidos através dos ouvidos". Já o termo identificação traz como definição: "reconhecer as características essenciais do som". A percepção engloba a identificação e refere-se a um nível cognitivo, já que implica na organização de dados sensoriais. A identificação está ligada ao reconhecimento, resultado de uma aprendizagem anterior. Qualquer estudo ou pesquisa sobre a percepção mínima deveria levar em consideração estas distinções.

- Parece que pequenas unidades são melhores identificadas quando se fraciona o inteiro. O fonema, neste caso, seria a unidade menor e a sílaba o inteiro com possibilidade de ser fracionado. Isto quer dizer que os estudos sobre a percepção mínima devem fazer a identificação do fonema isolado e do fonema inserido na sílaba para se ter uma identificação mais fiel.

- As experiências com seqüências silábicas indicam que primeiro há a representação mental do estímulo e em seguida é feita uma análise na tentativa de reproduzir o sinal

acústico. A detecção só começa realmente a acontecer quando há o encaixe entre a representação mental e a análise acústica. Isto sugere que exista no cérebro um padrão fisiológico para cada som em especial. Sugere também que as tentativas de identificar a percepção deverão passar pela reprodução do som a reconhecer, o que deverá exigir cuidados especiais, não só na seleção dos estímulos, como nas possibilidades de reprodução do ouvinte/cobaia.

4. CONCLUSÃO

Os estudos sobre a percepção mínima da fala feitos até o momento talvez não sejam conclusivos ou talvez nunca se chegue realmente a saber qual a unidade mínima da percepção da fala. No entanto, para os terapeutas da comunicação o que interessa é tomar conhecimento de que existem alterações da fala onde a sílaba é de maior importância e outros onde os fonemas têm um papel superior.

No tratamento dos distúrbios fonológicos, por onde começar o programa de reeducação? Alguns autores recomendam começar pela discriminação do som que é substituído, ou seja, levar o paciente reconhecer um /r/ ou /l/, mas se poderia começar também pela discriminação do som na sílaba /ri/ ou /li/. Nos distúrbios da voz, as características acústicas de certos fonemas demandam mais ou menos energia glótica ou subglótica, o que ajudaria no tratamento da voz rouca por excesso de tensão glótica ou na voz fraca por falta de tensão glótica. Nas afasias motoras e nas disartrias, a seqüência silábica é muito mais importante do que as características acústicas do fonema. Na deficiência de audição, o fonema isolado traz muito pouca informação tanto à percepção como na produção. Na dislexia e nos distúrbios de aprendizagem, existem correntes que ainda discutem quais seria o melhor método de alfabetização ou de reeducação: ensinar ou trabalhar com o fonema ou com a sílaba? Onde a criança percebe melhor? De que maneira ela integra melhor a informação?

O dilema de se trabalhar em nível da sílaba ou de fonema está longe de ser resolvido. Para termos respostas teremos de ter mais pesquisas voltadas para diferentes abordagens de tratamento e nas diversas características das patologias da fala tanto oral como escrita. Este artigo tem por objetivo levantar estas polêmicas e incentivar as pesquisas nesta área.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BORDEN, G. J. e HARRIS, K. S. (1980) *Speech Science Primer; Physiology, Acoustics and Perception of Speech*. Baltimore: Willians & Wilkins.

CHOMSKY, N. e HALLE, M. (1968) *The Sound Pattern of English*. New York: Harper and Row.

DELATTRE, P.; LIBERMAN, A. M. e COOPER, F. S. (1955) Acoustic loci and transitional cues for consonants. *Journal of the Acoustic Society of America*, vol. 27, 769-773.

FANT, G. (1967). *Models for the Perception of Speech and Visual Form*. MS: M.I.T. Press, ed. by W. Wathen-Dunn.

FOSS, D. S. e SWINNEY, D. A. (1973) On the Psychological Reality of the Phoneme: Perception, Identification and Consciousness. *Journal of Verbal Learning and Verbal Behavior*, vol. 15, 246-257.

FOSS, D. S. e BLANK, M. A. (1980). Identifying the Speech Code. *Cognitive Psychology*, vol. 12, 1-31.

ISSLER S. (2006) *Articulação e Linguagem*. Rio de Janeiro, Revinter, 4ª edição.

JAKOBSON, R. e HALLE, M. (1956) *Fundamentals of Language*. The Hague: Mouton.

JERGER, J. (1963). *Modern Developments in Audiology*. New York: Academic Press.

KNOBEL M. A física da fala e da audição. <http://www.ifi.unicamp.br/~knobel/f105/>.

LIBERMAN, A. M. (1970) The Grammars of Speech and Language. *Cognitive Psychology*, vol. 1, 301-323.

LIBERMAN, A. M.; SHANKWEILLER, D.; FISCHER, F.W. e CARTER, B. (1974) Reading and the Awareness of Linguistic Segments. *Journal of Experimental Child Psychology*, vol. 18, 199-208.

LIBERMAN, A. M.; COOPER, F., S. e SHANKWEILLER, D. P. (1967) Perception of the Speech Code. *Psychol.*, vol. 74, 431-461.

LIBERMAN, A. M. (1957) Some Results of Research on Speech Perception, *J. Acoustic Soc. Am.*, vol. 29, 117-123.

MASSARO, M. W. (1974) Perceptual Images Processing Time and Perceptual Units in Auditory Perception. *Psychological Rev.*, vol. 79, pg. 124-145.

MCNEILL, D. e LINDIG, K. (1973) The Perceptual Reality of Phonemes, Syllables, Words and Sentences. *Journal of Verbal Learning and Verbal Behavior*, vol. 2, 419-430.

MEHLER J.; DOMMERQUES J. Y. e FRAUENFELDER U. (1981) The Syllable's Role in Speech Perception. *Journal of Verbal Learning and Verbal Behavior*, vol. 20, 298-305.

MILLS, D. B. (1980) Effects of the Match between Listener Expectancies and Coarticulatory Cues on the Perception of Speech. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 6, 528-535.

MORAIS, J.; CARY, L.; ALEGRIA, J. E BERTELSON, P. (1979) Does Awareness of Speech as a Sequence of Phones Arise Spontaneously? *Cognition*, vol. 7, 323-331.

NICOLOSO, L.; HARRYMAN, E. e KRESHECK J. (1996). *Vocabulário dos Distúrbios da Comunicação: Fala, Linguagem e Audição*. Porto Alegre: Artes Médicas.

SAVIN, H. B., e BEVER, T. G. (1970) The Non Perceptual Reality of the Phoneme. *Journal of Learning and Verbal Behavior*, vol. 9, 295-303.

REGINA JAKUBOVICZ

FONOAUDIÓLOGA, PROFESSORA TITULAR DA UNIVERSIDADE ESTÁCIO DE SÁ - RIO DE JANEIRO (BRASIL).