

Características prosódicas da fala do intérprete e sua influência na percepção de qualidade da interpretação

Layla Penha*

Sabemos da importância da qualidade do trabalho do intérprete, mas qual é a definição de qualidade? Tradicionalmente, embora se considere desejável que o intérprete produza uma fala expressiva, as questões prosódicas assumem um papel secundário na definição e avaliação da qualidade do trabalho do intérprete. Estudos realizados entre intérpretes e ouvintes usuários da interpretação mostram que os aspectos prosódicos são considerados menos relevantes na avaliação dos itens que contribuem para a qualidade da interpretação e, em uma lista de importância, aparecem atrás dos critérios linguísticos, mais atrelados à consistência de sentido, coesão lógica e outros (Bühler, 1986; Kurz, 1989, Moser 1996; Chiaro e Nocella, 2004; Pöchhacker e Zwischenberger, 2010). No entanto, quando os mesmos avaliadores têm como tarefa analisar a produção oral dos intérpretes, tendem a atribuir maior qualidade e mais alto nível de compreensibilidade àquelas versões que são superiores nas mesmas características prosódicas consideradas menores na avaliação da qualidade do trabalho do intérprete, como voz agradável, vivacidade e fluência, sugerindo que talvez os elementos prosódicos sejam igualmente

* Layla Penha é intérprete de conferências, mestre em Linguística Aplicada pela PUC-SP e professora do Curso Sequencial de Interpretação de Conferências da PUC-SP. Este artigo se baseia em sua dissertação de Mestrado.

importantes na avaliação de qualidade e compreensibilidade da interpretação.

Para tentarmos entender como essa percepção de qualidade acontece na prática, desenvolvemos uma pesquisa de avaliação da fala interpretada. Os resultados sugerem que o trabalho do intérprete não pode ser resumido ao conteúdo estritamente linguístico da fala. Mensagens importantes são passadas pela forma em que nos comunicamos através de elementos prosódicos, e essa forma tem tanto peso quanto o conteúdo do que dizemos, como veremos a seguir.

1. Participantes da Pesquisa

O estudo foi desenvolvido com doze intérpretes profissionais do sexo feminino, todas falantes nativas de português brasileiro (PB) e com afiliação em uma ou mais entidades profissionais, nomeadamente Associação Internacional de Intérpretes de Conferência (AIIC), Associação Profissional de Intérpretes de Conferência (APIC) ou *American Translators Association* (ATA). O tempo médio de exercício da profissão das intérpretes na época da pesquisa era de 15 anos e sua idade, entre 41 e 53 anos.

Extratos da produção oral das intérpretes foram submetidos à avaliação de um grupo de 90 juízes, todos recém-formados ou cursando Tradução ou Interpretação na Pontifícia Universidade Católica de São Paulo, com idades entre 18 e 60 anos.

As transcrições dos mesmos extratos de fala também foram submetidas à avaliação de três profissionais com experiência na formação de intérpretes. Os avaliadores eram intérpretes profissionais, membros das referidas associações AIIC, APIC e ATA, com idade entre 45 e 51 anos, tempo médio de profissão de 11 a 15 anos e experiência na formação de intérpretes de 3 a 12 anos.

2. Procedimentos metodológicos

2.1 Seleção e descrição da fala a ser interpretada

A fala utilizada para o exercício de interpretação foi uma palestra intitulada *Don't Insist on English*, proferida por Patricia Ryan, professora da Universidade de Zayed, em Dubai. A palestra foi filmada em dezembro de

2010 no evento TEDxDubai e está disponível para download no acervo do TED TALKS¹. A duração total da palestra é de 10min35s.

2.2 Coleta do corpus

As intérpretes foram audiogravadas através do equipamento Sanako Lab 100, disponibilizado pela Associação Alumni, em São Paulo, capital, em quatro sessões consecutivas no dia 19 de agosto de 2013. Cada profissional trouxe seu próprio *laptop* e fone de ouvido para a sessão de gravação para que a versão a ser interpretada fosse baixada em seu equipamento. Foram utilizadas entre três e quatro cabines de interpretação em cada sessão, com ocupação de uma intérprete por cabine.

Nas cabines, as intérpretes foram instruídas a posicionar o microfone da cabine em volta do pescoço, a aproximadamente 7cm de distância da boca, para sua primeira tarefa: falar sobre um tema de sua livre escolha por aproximadamente um minuto e meio (fala semiespontânea). O objetivo dessa primeira tarefa era ter um registro inicial da fala de cada profissional fora da tarefa de interpretação, para posterior análise e comparação com a fala interpretada em termos de taxa de elocução (velocidade da fala), distribuição e número de pausas e grupos entoacionais (sequências distintivas de tons numa elocução que compõem uma unidade de sentido), e frequência fundamental - f_0 (que corresponde, de modo subjetivo, à nossa percepção de variações de altura e de intensidade na fala).

Na sequência, as intérpretes iniciaram a gravação da interpretação. Através de seus fones de ouvido conectados aos *laptops*, as profissionais tinham acesso ao áudio e ao vídeo da palestra, o que reproduzia de maneira bastante fidedigna um ambiente típico de trabalho, pois podiam acompanhar a palestrante visualmente durante a palestra. Mais uma vez, a gravação da interpretação foi feita através do microfone da cabine, repetindo o ajuste anterior. Como as intérpretes utilizaram seus próprios fones de ouvido como fonte de entrada de áudio e os microfones na cabine

¹ <http://www.ted.com/search?q=patricia+ryan>

como fonte de saída de áudio, não houve interferência do áudio de partida no canal de gravação.

A gravação foi feita na extensão .mp3 e posteriormente convertida para .wav, canal mono, para análise de fala pelo Praat, um *software* livre desenvolvido por Paul Boersma e David Weenink, do Instituto de Ciências Fonéticas de Amsterdam² (Boersma, 2002).

2.3 Análise acústica dos extratos de áudio

Findo o trabalho de coleta, os áudios completos de cada intérprete foram abertos no software Praat. Através das ferramentas de edição, o extrato de fala semiespontânea foi separado da fala interpretada e analisado. Na fala interpretada, três extratos representativos do discurso foram selecionados para análise e avaliação – um inicial, na abertura da palestra (com tempo de fala total, descontados os períodos de aplauso, de 34s), um medial, produzido no meio da palestra, a partir do minuto 5:42 (tempo de fala total, 34s), e um final, na conclusão da palestra (tempo de fala total, 25s). Um segmento adicional adjacente ao segmento medial produzido por 3 intérpretes aleatoriamente selecionadas foi também incorporado na tarefa de avaliação perceptiva de todos os grupos, com o objetivo de verificarmos a uniformidade das respostas.

Dentro do segmento inicial, havia dois períodos de aplauso no discurso de partida, com duração de aproximadamente 5,5s e 2,5s respectivamente. Na fala das intérpretes, esse tempo de espera se traduzia em pausas silenciosas, mas com durações diferentes, dependendo da distância de cada intérprete em relação à fala de partida. Para deixar as gravações mais uniformes, optou-se por editar as pausas silenciosas nesses dois pontos, eliminando qualquer período superior a 1 segundo. O mesmo foi feito com a fala de partida em inglês, para maior homogeneidade na comparação dos resultados. Os outros extratos não apresentaram dificuldades e não foram editados.

² Disponível para *download* no site www.praat.org.

Os segmentos extraídos da fala de cada intérprete, juntamente com os segmentos de partida em inglês, foram analisados no software Praat. No total, 55 segmentos foram analisados: 12 segmentos contendo a fala semiespontânea das intérpretes, 4 segmentos de partida (inicial, medial, final e adicional), 12 segmentos contendo os extratos iniciais de cada interpretação, 12 segmentos contendo os extratos mediais, 12 segmentos contendo os extratos finais e 3 segmentos contendo o extrato adicional. Os extratos foram analisados quanto à quantidade e à distribuição de grupos entoacionais (GE), quantidade e distribuição de pausas (P), e quantidade e distribuição de alongamentos vocálicos (AV). Também foram tomadas as medidas de frequência fundamental (em hertz - Hz) média, mínima e máxima de cada extrato.

A Figura 1 a seguir mostra como cada segmento foi analisado acusticamente. As diferentes camadas mostram a divisão em grupos entoacionais (GE), incidência de pausas (P) e alongamentos vocálicos (AV). A segmentação desses elementos foi realizada com base no espectrograma de banda larga com referência à forma de onda. O espectrograma de banda larga é um gráfico que apresenta a frequência na ordenada, o tempo na abscissa e a intensidade no contraste de cor (quando o espectrograma é apresentado em tons de cinza, quanto mais escura a região de frequência, maior a intensidade). Os números no eixo vertical do espectrograma de banda larga mostram a faixa de frequência em Hz.

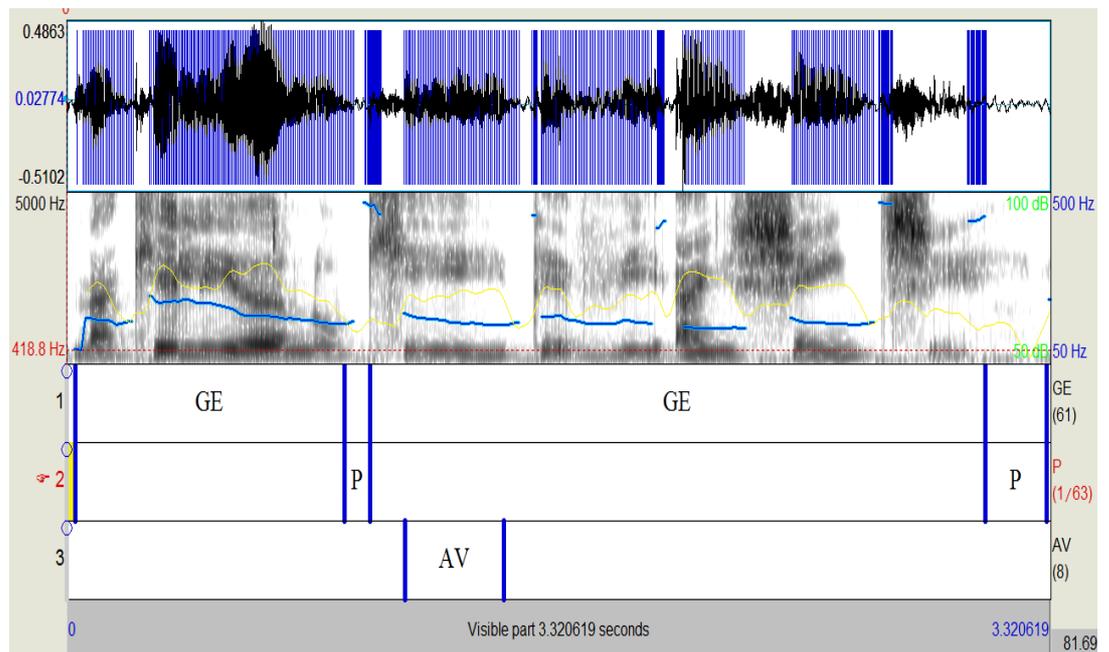


Figura 1 – Extrato de segmento analisado

2.4 Avaliação perceptiva da qualidade da produção oral dos intérpretes

Os extratos foram então submetidos à avaliação de 90 juízes. Os juízes foram divididos em três grupos para analisar 15 extratos de fala, com cada grupo responsável por analisar 12 extratos de fala diferentes (4 iniciais, 4 mediais e 4 finais). Como mencionado, 3 extratos adicionais foram incorporados em todas as versões para verificação de uniformidade no comportamento de avaliação dos juízes de cada grupo. Cada versão trazia uma produção de cada uma das intérpretes, de maneira que todos os 90 juízes avaliaram pelo menos uma fala de cada intérprete.

A tarefa dos juízes era responder um questionário baseado na Escala de Likert para verificação de avaliação de qualidade e grau de compreensibilidade. Esse método de avaliação mede uma resposta positiva ou negativa a um determinado parâmetro de avaliação e é geralmente aplicado em uma escala de cinco a sete pontos. Nesse estudo, a tarefa dos respondentes era especificar seu nível de concordância com cada afirmação, dentro de uma escala de cinco pontos onde 1 significava que o extrato de áudio não atendia em nada o quesito avaliado, e 5 significava

que o extrato atendia extremamente o quesito avaliado. No total, sete parâmetros foram avaliados, a saber: Agradabilidade (nada agradável/extremamente agradável), Credibilidade (nada confiável/extremamente confiável), Naturalidade (nada natural/extremamente natural), Fluidez (nada fluida/extremamente fluida), Segurança (nada segura/extremamente segura), Profissionalismo (nada profissional/extremamente profissional) e Compreensibilidade (nada fácil de entender/extremamente fácil de entender).

A plataforma utilizada foi a Survey Gizmo³, uma ferramenta de pesquisas online. Cada juiz recebeu um link via e-mail ou Facebook para identificação e ativação do questionário. Somente os juízes convidados poderiam responder as perguntas. O questionário foi montado de forma a considerar válidas apenas as respostas completas. Todas as perguntas eram mandatórias, ou seja, todos os resultados válidos contemplavam 100% das perguntas respondidas.

2.5 Avaliação da qualidade da interpretação em relação ao discurso interpretado em termos de conteúdo semântico-discursivo

A transcrição das falas também foi submetida à avaliação de conteúdo por três profissionais com experiência na formação de intérpretes, já apresentados anteriormente, cuja tarefa era analisar cada extrato isoladamente em comparação ao original. Consoante ao estudo realizado por Gile (1999), os avaliadores deveriam se concentrar no número de erros (E), número de omissões (O) e manutenção do sentido original (S) para a avaliação da qualidade da interpretação, atribuindo uma nota final (T) de 0 a 5 (onde 0 seria considerado insuficiente e 5 excelente) a cada extrato. As notas de cada intérprete por extrato foram somadas, e cada intérprete recebeu uma pontuação final cumulativa de 0 a 15. Essa pontuação final foi então dividida por três para termos uma média final possível entre 0 e 5 para cada intérprete.

³ Disponível em <http://www.surveygizmo.com/>

2.6 Extratos de partida do discurso em inglês

Descontados os intervalos para aplauso, especialmente contidos no segmento inicial da palestra, os quatro extratos de partida do discurso utilizados nesta pesquisa totalizaram respectivamente 2,49 palavras por segundo, 2,53 palavras por segundo, 1,83 palavras por segundo e 2,31 palavras por segundo. Se convertermos esses números para palavras por minuto, teremos 149,4 palavras por minuto para o segmento inicial, 151,8 para o segmento medial, 109,8 para o segmento final e 138,72 para o segmento adicional. De acordo com a tabela desenvolvida por Tauroza & Allison (1990, p. 102) para o inglês britânico, essa média revela uma taxa de elocução considerada de normal a moderadamente lenta para fala em palestra. A frequência média observada nos segmentos inicial, medial, final e adicional foi de 219,32 Hz, com a média de frequência mais alta sendo verificada no primeiro segmento (239,89 hz).

3. Resultados

3.1 Fala Semiespontânea

A fala semiespontânea das intérpretes foi coletada antes da fala interpretada em um contínuo de gravação. Finalizada a coleta dos áudios, os arquivos sonoros foram abertos no software Praat para inspeção das características acústicas, análise e medição em cada trecho de fala da duração de pausas silenciosas, grupos entoacionais, alongamentos vocálicos e das produções de fala em milissegundos e dos valores máximos e mínimos de f_0 em Hz. Uma primeira janela foi aberta para podermos visualizar os dois momentos da gravação (vide Figura 2). O primeiro bloco de ondas sonoras indica a fala semiespontânea e o segundo bloco, a fala interpretada.

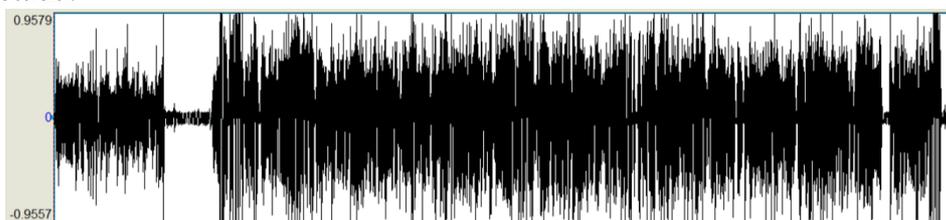


Figura 2 – Forma de onda do corpus de gravação completo

Como podemos observar na figura acima a partir das diferenças de amplitude das ondas sonoras, algumas intérpretes variaram a intensidade entre a fala semiespontânea e a fala interpretada. Perceptivamente, também observamos essa discrepância em termos de altura e intensidade de voz. Como a gravação de ambas as falas foi realizada em um contínuo, dentro dos mesmos ajustes de gravação, podemos apontar a diferença na amplitude da forma da onda entre os dois momentos de fala e as diferenças auditivas percebidas. Com exceção de uma intérprete (Sujeito de pesquisa 9 ou S9) todas as variações foram para maior na fala interpretada.

A fala semiespontânea foi então isolada do todo para análise, levando em consideração o número de grupos entoacionais, o número e a distribuição de pausas silenciosas, respiratórias ou vocais (através da utilização de “ums”, “ahns” e outros), a presença de alongamentos vocálicos, a taxa de elocução e os valores de frequência fundamental (mínimos e máximos), como podemos observar na Figura 3.

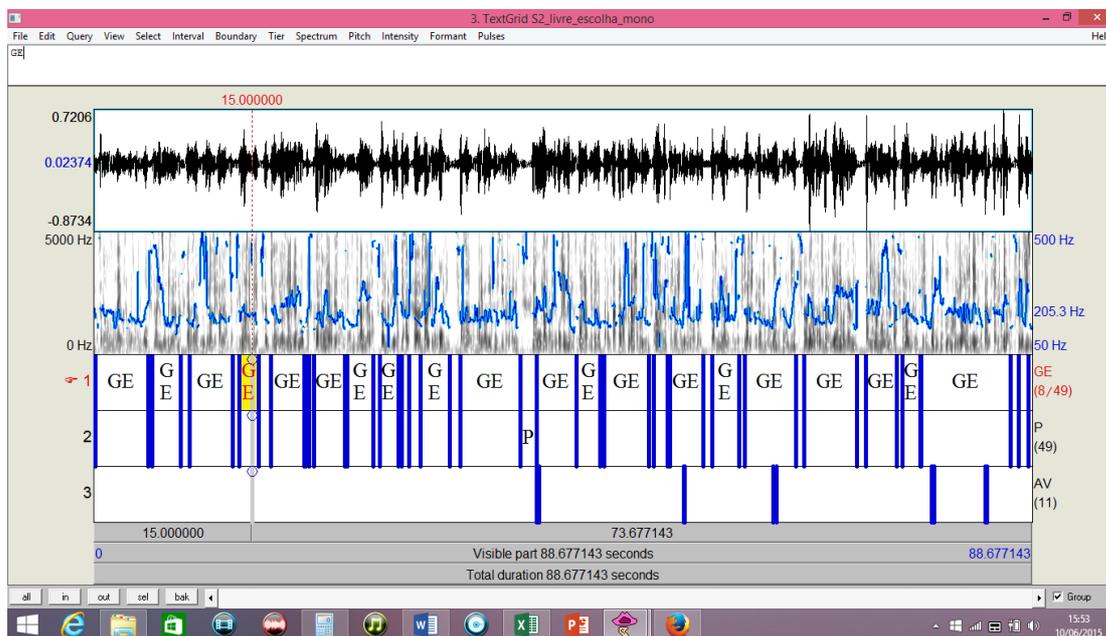


Figura 3 – Forma de onda, espectrograma de banda larga, camada de divisão dos grupos entoacionais (GE), camada com delimitações de pausas silenciosas e camada com delimitações de alongamentos vocálicos (AV) na fala semiespontânea.

A fala semiespontânea teve duração de 83,52 a 106,87 segundos. As intérpretes apresentaram frequência fundamental média nos segmentos entre 175,68 e 237,84 Hz, consoante aos dados encontrados na literatura que indicam uma frequência fundamental média de 204 Hz (Baraldi, 2007) para mulheres adultas (entre 18 e 45 anos) falantes do português brasileiro. As frequências fundamentais mínimas e máximas foram medidas posicionando o cursor nos valores mínimo e máximo de f0 em cada trecho de fala no traçado gerado por meio da função *show pitch* do Praat. Erros de registro de f0 foram descartados. Com as medidas de frequência fundamental mínima e máxima coletadas para cada segmento, as configurações de *pitch* (*pitch settings*) foram alteradas, de maneira que o campo de frequência a ser computado ficasse entre os valores mínimo e máximo identificados para cada extrato. A partir daí, utilizamos a função *get pitch* para obter a frequência fundamental média de cada segmento. O procedimento foi repetido para cada extrato analisado (inicial, medial, final e adicional de partida e interpretado).

A taxa de elocução na fala de livre escolha (semiespontânea) variou de 1,81 a 2,92 palavras por segundo, valores correspondentes a uma variação de 108,6 a 175,2 palavras por minuto, achados também consoantes com a literatura (Zackiewics & Andrade, 2000).

A frequência de pausas foi de uma pausa a cada 2,52 segundos a uma pausa a cada 3,87 segundos, com número de pausas total entre 22 e 36. No geral, a distribuição de pausas acompanhou a distribuição dos grupos entoacionais (de 23 a 37 grupos entoacionais em cada trecho de fala). Pausas de hesitação no meio de um grupo entoacional foram achados infrequentes.

3.2 Fala interpretada

A taxa de elocução das falas interpretadas apresentou grande variabilidade, com algumas intérpretes se aproximando mais da taxa de elocução de partida e outras utilizando taxas mais lentas ou mais rápidas, sem um padrão estabelecido. No entanto, entre as 12 intérpretes gravadas, dez diminuíram sua taxa de elocução ao iniciar a interpretação, mantendo

no geral uma taxa consistentemente inferior à taxa da fala semiespontânea, independentemente da taxa de elocução da fala de partida.

Em relação à distribuição e número de grupos entoacionais e pausas, as falas interpretadas apresentaram mais frequentemente um número de pausas superior ao número de grupos entoacionais, indicando a utilização de pausas silenciosas, respiratórias ou vocais (através da utilização de “ums”, “ahns” e outros) não só entre grupos entoacionais, mas também intragrupos (ou seja, não somente ao término de uma seqüência de sons correspondente a uma unidade de sentido, mas no meio de uma unidade de sentido).

Todas as intérpretes registraram aumento na frequência fundamental média na fala interpretada em comparação à fala semiespontânea. No geral, o segmento inicial foi o que apresentou os maiores picos de frequência, registrando variação entre 194,66 e 267,95 Hz, enquanto na fala semiespontânea os valores variaram entre 175,68 e 237,84 Hz. É interessante observar que o segmento inicial foi também o segmento com registro de maior frequência na fala de partida, o que pode ser explicado por um maior nervosismo em início de trabalho, tanto para intérpretes quanto para palestrantes.

3.3 Avaliação perceptiva da qualidade da produção oral dos intérpretes

Como já mencionado, noventa juízes participaram da pesquisa, com um total de 21 juízes da área de tradução (23,3%) e 69 da área de interpretação (76,5%). Vinte e cinco juízes (27,78%) haviam concluído seu curso nos últimos dois anos, com o restante de 65 juízes (72,2%) ainda matriculados. Os juízes foram escolhidos por afinidade com o discurso de partida, sendo assim uma “plateia ideal” para o trabalho dos intérpretes. A palestra de Patricia Ryan defende a tradução e o compartilhamento de ideias, e o quão importante é não substituímos nossos idiomas nativos pelo uso do inglês.

Os juízes foram divididos em grupos de trinta para a avaliação das três versões. Cada versão de avaliação possuía 15 segmentos ao todo – 4 iniciais, 4 mediais, 4 finais e 3 adicionais, de forma que cada juiz tivesse a oportunidade de ouvir cada intérprete pelo menos uma vez. Os três

segmentos adicionais foram repetidos em todas as versões para verificação da uniformidade no julgamento das falas dos intérpretes, como podemos verificar na tabela abaixo:

V1	V2	V3
S7 adicional	S2 inicial	S12 adicional
S1 inicial	S5 inicial	S3 inicial
S4 inicial	S9 inicial	S6 inicial
S7 inicial	S12 inicial	S10 inicial
S11 inicial	S5 adicional	S8 inicial
S5 adicional	S8 medial	S7 adicional
S12 medial	S10 medial	S11 medial
S9 medial	S6 medial	S7 medial
S5 medial	S3 medial	S4 medial
S2 medial	S12 adicional	S1 medial
S12 adicional	S7 final	S5 adicional
S10 final	S1 final	S9 final
S3 final	S11 final	S2 final
S8 final	S4 final	S12 final
S6 final	S7 adicional	S5 final

Tabela 1 - Distribuição dos extratos de fala em cada versão de avaliação

Cada extrato foi avaliado através de uma Escala de Likert de 5 pontos. A tarefa dos juízes era a de escolher, para cada extrato ouvido, se o áudio não atendia o quesito em nada, se o atendia pouco, se não o atendia nem muito nem pouco, se o atendia muito ou se o atendia extremamente. Os resultados foram então ponderados para atribuição de nota, com “Nada” recebendo peso 1, “Pouco” recebendo peso 2, “Nem Muito nem Pouco”, peso 3, “Muito”, peso 4 e “Extremamente”, peso 5. As notas atribuídas a cada extrato foram então multiplicadas por seu peso ponderado, somadas e divididas pelo número de juízes, para chegarmos a uma média final por extrato.

Apesar de diferenças em números absolutos, os três grupos chegaram a um mesmo resultado em termos de classificação dos extratos adicionais, com S5 recebendo as maiores pontuações, vindo S12 em segundo lugar e S7 em terceiro lugar, o que sugere uniformidade de avaliação.

A média final de cada intérprete na avaliação da fala interpretada pode ser encontrada na Tabela 2. Os quesitos em que as intérpretes apresentaram as menores notas foram Agradabilidade, Naturalidade, Fluidez e Segurança (nessa ordem).

Na média, nenhuma intérprete teve avaliação abaixo de 2 nem atingiu a marca de 5. As médias finais das intérpretes estão na última coluna à direita.

Média	Agrad	Conf	Nat	Flu	Seg	Prof	Compr	MF
S1	3,32	3,6	3,44	3,52	3,59	3,47	3,54	3,49
S2	3,96	4,07	3,9	4,04	4,11	4,12	4,04	4,03
S3	2,67	2,51	2,43	2,32	2,42	2,61	2,38	2,47
S4	3,81	3,8	3,7	3,77	3,74	3,74	3,9	3,78
S5	4,05	4,04	3,8	3,83	3,96	4	3,97	3,94
S6	3,55	3,55	3,54	3,48	3,44	3,5	3,67	3,52
S7	2,24	3,15	2,48	3,01	3,35	3,01	3,32	2,93

S8	2,1	2,93	2,7	2,68	3	2,97	3,08	2,79
S9	2,78	2,74	2,66	2,46	2,63	2,76	2,91	2,7
S10	2,24	3,1	2,78	3,18	3,46	3,23	3,16	3,02
S11	2,85	3,46	3,24	3,38	3,74	3,52	3,8	3,42
S12	3,18	3,22	3,15	3,14	3,12	3,17	3,34	3,19

Tabela 2 – Média ponderada final de cada quesito por Sujeito

3.4 Avaliação da qualidade da interpretação em relação ao discurso interpretado em termos de conteúdo semântico-discursivo

As médias foram calculadas a partir da soma das notas atribuídas a cada quesito da performance das intérpretes em cada um dos extratos de fala. Os resultados das avaliações de conteúdo de cada avaliador foram então comparados para verificação de uniformidade na ordem de classificação dos sujeitos. Quando as médias finais eram as mesmas, o primeiro critério de desempate utilizado foi contabilizar as maiores notas no nível dos segmentos, primeiramente considerando a incidência total da nota mais alta, depois a incidência total da segunda nota mais alta e assim por diante. Só foi considerado empate no caso de pontuações com exatamente a mesma composição de notas, o que aconteceu duas vezes. Nesse caso, a posição imediatamente abaixo ficou vaga. A tabela abaixo mostra a classificação dos Sujeitos de acordo com cada Avaliador.

	Avaliador 1	Avaliador 2		Avaliador 3
1º lugar	S2	S5	S2	S5
2º lugar	S1			S1
3º lugar	S10	S10		S2
4º lugar	S4	S1		S7
5º lugar	S7	S7		S11
6º lugar	S5	S12		S9 S10
7º lugar	S12	S8		

8º lugar	S11	S11	S4
9º lugar	S9	S6	S8
10º lugar	S8	S4	S6
11º lugar	S6	S9	S12
12º lugar	S3	S3	S3

Tabela 3 – Classificação final de cada Sujeito por avaliador

Cinco sujeitos – S1, S2, S5, S7 e S10 – apareceram nas primeiras seis posições nas avaliações dos três juízes. Três sujeitos – S3, S6 e S8 – apareceram nas seis últimas posições. Os sujeitos S4, S9, S11 e S12 aparecem uma vez entre os seis primeiros e duas vezes entre os seis últimos colocados.

A Tabela abaixo traz a média e classificação final de cada Sujeito.

	Avaliador	Avaliador	Avaliador	MÉDIA	Classif.
	1	2	3		
S1	4	4,83	4,67	4,5	3º
S2	4,33	4,93	4,67	4,64	1º
S3	1,00	1,83	2,33	1,72	12º
S4	3,67	4,53	4,33	4,17	6º
S5	3,67	4,93	5,00	4,53	2º
S6	2,00	4,53	3,67	3,4	11º
S7	3,67	4,77	4,67	4,37	4º
S8	2,33	4,60	4,00	3,64	10º
S9	2,67	4,43	4,33	3,81	9º
S10	3,67	4,87	4,33	4,29	5º
S11	3,00	4,60	4,67	4,09	7º
S12	3,33	4,77	3,67	3,92	8º

Tabela 4 – Nota final atribuída a cada sujeito por avaliador

3.5 Comparação das análises perceptiva e de conteúdo semântico-discursivo

De um modo geral, as médias finais de avaliação do conteúdo da fala interpretada mostraram um espectro de pontuação mais amplo que o da análise perceptiva, com a média mais baixa sendo 1,72 e a mais alta 4,64 para a primeira, contra 2,47 e 4,03 para a segunda. Abaixo podemos ver uma comparação das avaliações. As diferentes cores sinalizam a classificação das intérpretes. Em azul, resultados iguais ou com apenas uma posição de diferença. Em verde, duas posições de diferença. Em vermelho, três posições de diferença e, em destaque, os dois sujeitos que apresentaram diferença de cinco posições.

Intérprete	Média Transcrições	Classificação	Média Arquivos Sonoros	Classificação
S1	4,5	3º	3,49	5º
S2	4,6433	1º	4,03	1º
S3	1,72	12º	2,47	12º
S4	4,1767	6º	3,78	3º
S5	4,5333	2º	3,94	2º
S6	3,4	11º	3,52	4º
S7	4,37	4º	2,93	9º
S8	3,6433	10º	2,79	10º
S9	3,81	9º	2,7	11º
S10	4,29	5º	3,02	8º
S11	4,09	7º	3,42	6º
S12	3,9233	8º	3,19	7º

Tabela 5 - Comparação das médias finais ponderadas das avaliações das transcrições e dos arquivos sonoros.

Seis sujeitos da pesquisa tiveram exatamente a mesma classificação nas duas avaliações: S2 – 1ª posição; S5 – 2ª posição; S11 – 7ª posição; S12 – 8ª posição; S8 – 10ª posição e S3 – 12ª posição.

Os Sujeitos 2 e 5 receberam as melhores pontuações em ambas as avaliações. Se observarmos a análise acústica da produção da interpretação, veremos que as duas intérpretes apresentaram pouca variabilidade entre a produção semiespontânea e a produção interpretada em relação à intensidade, número e distribuição de pausas e grupos entoacionais e alongamentos vocálicos. As variações de taxa de elocução e de frequência média foram inferiores a 20%, sendo a primeira a menor na fala interpretada e a segunda a maior na fala interpretada. Os quesitos com maior pontuação na análise perceptiva foram o de Profissionalismo (S2) e Agradabilidade da voz (S5). Os quesitos mais adversos foram Naturalidade (S2) e Fluidez (S5).

Os Sujeitos 11, 12, 8 e 3 aparecem entre os seis piores classificados em ambas as avaliações. Os Sujeitos 8 e 3 apresentam a menor taxa de elocução do grupo, com produção média de taxa de elocução na fala interpretada de 117,2 palavras por minuto e 103 palavras por minuto, respectivamente, enquanto o restante das intérpretes teve média de 133,88 e o discurso de partida teve 137,4, o que significa uma diferença de aproximadamente 20%. O S3 tem Fluidez com a pior pontuação e Agradabilidade da voz como o quesito melhor avaliado, enquanto para o S8 este é o parâmetro com pior avaliação. Compreensibilidade é a melhor pontuação do S8.

O S12 apresenta a maior incidência de pausas interrompendo unidades de sentido (grupos entoacionais) e a maior frequência fundamental média. O S11 apresenta a menor frequência fundamental média. Os quesitos com pior avaliação foram Segurança (S12) e Agradabilidade de voz (S11). O mais bem avaliado foi Compreensibilidade para ambos, o que não lhes garante percepção de qualidade.

Os Sujeitos 1 e 9 apresentaram diferenças de duas posições entre a avaliação semântica-discursiva e a avaliação perceptiva da fala, a menor para a segunda, com o S1 passando da 3ª para a 5ª posição e o S9, da 9ª para a 11ª posição.

O S4 e o S10 registraram uma diferença de 3 posições cada, com o S4 passando da 6ª para a 3ª posição e o S10 da 5ª para a 8ª posição. As melhores avaliações ficaram em Compreensibilidade (S4) e Segurança (S10) e as piores, em Naturalidade (S4) e Agradabilidade de voz (S10).

Agradabilidade de voz foi o quesito com menor pontuação para 5 das 12 intérpretes e com a maior pontuação para duas delas, talvez sugerindo que a qualidade vocal da intérprete pode ter um papel importante na percepção positiva de seu trabalho. No entanto, a Agradabilidade da voz não apresentou impacto na avaliação de Compreensibilidade da interpretação, uma vez que esses dois descritores tiveram, no geral, avaliação diretamente oposta, com a Agradabilidade de voz tendo a pior avaliação no cômputo de todas as intérpretes (média geral de 3,06) e Compreensibilidade, a melhor (média geral de 3,42).

Efetivamente, a maior diferença encontrada entre as avaliações ficou com os Sujeitos 6 e 7. O S6 passou da 11ª posição em conteúdo para a 4ª posição na análise perceptiva da fala. Em movimento contrário, o S7 passou da 4ª posição em conteúdo para a 9ª posição na análise perceptiva. Enquanto o S6 registra na análise acústica padrão semelhante aos sujeitos melhor avaliados na análise perceptiva (pouca variabilidade entre a produção semiespontânea e a produção interpretada em relação à intensidade, número e distribuição de pausas e grupos entoacionais e alongamentos vocálicos, taxa de elocução e de frequência média com diferença inferior a 20%), o S7 apresenta a maior variabilidade entre falas, com grande aumento de intensidade e aumento de frequência fundamental superior a 30%. O pior quesito para o S7 foi Agradabilidade de voz, e o quesito melhor avaliado foi Segurança. A percepção de segurança está relacionada na literatura sobre expressividade com um f_0 mais baixo. O S7 está no grupo dos sujeitos que têm o f_0 abaixado.

Dada a destacada importância da compreensibilidade na tarefa de interpretação, consideraremos, a seguir, as porcentagens obtidas na avaliação desse descritor em grau alto (muito e extremamente) nos extratos de fala analisados e discutiremos o impacto dos fatores prosódicos em relação aos resultados obtidos. A compreensibilidade baixa remete ao

“nada ou pouco” compreensível e a compreensibilidade média, ao “nem muito nem pouco” no teste de diferencial semântico aplicado.

A Tabela 6 apresenta os resultados da avaliação da fala de cada intérprete em relação aos três trechos (inicial, medial e final) de fala interpretada.

Sujeito	Compreensibilidade	Inicial	Medial	Final
S1	Alta	46,70%	70,00%	46,60%
S2	Alta	60,00%	86,70%	90,00%
S3	Alta	23,30%	10,00%	10,00%
S4	Alta	76,70%	70,00%	76,70%
S5	Alta	60,00%	76,60%	93,30%
S6	Alta	60,00%	73,40%	63,30%
S7	Alta	50,00%	43,40%	43,40%
S8	Alta	40,00%	16,70%	26,70%
S9	Alta	20,00%	36,70%	20,00%
S10	Alta	60,00%	20,00%	33,30%
S11	Alta	43,30%	90,00%	60,00%
S12	Alta	33,30%	30,00%	70,00%

Tabela 6 – Compreensibilidade alta atribuída a cada Sujeito por extrato

Como podemos observar, somente os sujeitos 2, 4, 5 e 6 alcançaram porcentagens de compreensibilidade acima de 60% na avaliação dos três trechos de fala interpretada, o que nos parece um valor baixo se comparado à avaliação do conteúdo semântico da fala interpretada.

Os sujeitos 1, 7, 10 e 12 foram avaliados com porcentagens abaixo de 50% em dois dos extratos e os sujeitos 3, 8 e 9, nos 3 extratos. Vale lembrar que os sujeitos 1 e 7 tiveram média na avaliação das transcrições superior a

4 ($S1 = 4,5$ e $S7 = 4,37$), ou seja, o conteúdo estava presente em suas falas, mas a maneira de expressá-lo trouxe dificuldades de compreensão. Esses resultados sugerem que a compreensibilidade da fala do intérprete possa ser afetada pelas características prosódicas.

4. Conclusões

De uma maneira geral, o trabalho nos trouxe elementos importantes para a pesquisa do trabalho de interpretação simultânea. Pudemos ver pelos resultados que a fala interpretada para o português brasileiro mostra padrão muito similar às falas interpretadas para o inglês e o hebraico estudadas por Schlesinger (1994). Também no PB, foram encontradas pausas no meio de estruturas gramaticais, uso característico das proeminências, com sílabas “inesperadas” carregando o *pitch accent*, ou valor de *pitch* máximo, marcação do *pitch* final de fronteira, com tendência de os intérpretes concluírem suas frases com um *pitch* elevado, não concludente, e produção sem ritmo regular, com alterações fora do padrão tanto em duração quanto em taxa de elocução.

Isso demonstra que a fala interpretada é uma fala característica por si só, diferente de outras modalidades de fala, seja ela espontânea, semiespontânea, fala em leitura ou fala em *shadowing*. Este artigo comprova que os dois estilos analisados, o semiespontâneo e o interpretado, são estilos diferenciados. Por essa razão, além de compararmos a avaliação da fala interpretada com outras modalidades de fala, como tem sido feito nos estudos mais recentes de qualidade na interpretação (Holub, 2010; Rennert, 2010; Lenglet, 2015), é também fundamental entendermos como a prosódia afeta a avaliação de qualidade da interpretação.

É justamente isso que tentamos desenvolver neste trabalho, e os resultados demonstram, em termos de classificação geral das intérpretes, haver certa compatibilidade entre a avaliação da transcrição da produção dos sujeitos e a avaliação das características da fala. No entanto, equivalência em termos de sentido não é garantia de uma melhor ou pior avaliação da qualidade da interpretação. Além de 10 das 12 intérpretes registrarem notas consistentemente mais baixas em produção oral, vimos

que a própria ordem de classificação geral das intérpretes também foi alterada.

Com base nos resultados, podemos inferir que as intérpretes que mais aproximaram as características prosódicas da fala interpretada de sua fala semiespontânea tenderam a receber uma melhor avaliação de qualidade e compreensibilidade da interpretação. Aquelas que mais se afastaram das características prosódicas identificadas na fala semiespontânea tenderam a ser pior avaliadas. No entanto, ainda mais estudos serão necessários para revalidarmos essa afirmação.

Ainda precisamos caminhar muito na compreensão de como a fala interpretada pode ser avaliada, mas aprendemos através deste estudo que as questões prosódicas têm papel importante na avaliação de qualidade e compreensibilidade da fala interpretada. Como vimos na Tabela 6, das doze intérpretes aqui avaliadas, apenas 3 atingiram mais de 60% em alta compreensibilidade. Isso é muito pouco. Precisamos utilizar nosso conhecimento para incluir nos programas de formação para intérpretes o estudo das questões prosódicas, tão fundamentais para nossa profissão.

Referências

- BARALDI, G. et al. Estudo da frequência fundamental da voz de idosas portadoras de diferentes graus de perda auditiva. **Revista Brasileira de Otorrinolaringologia**, v.73, n.3, p. 378-383, 2007, disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992007000300013&lang=pt> Acesso em: 7 nov. 2017.
- BOERSMA, P; WEENINK, D. Praat. **Doing phonetics by computer**. Amsterdam, 2002, disponível em: <<http://www.fon.hum.uva.nl/praat/>>.
- BUHLER, H. Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. **Multilingua**, 5 (4): 231-235, 1986.
- CHIARO, D. & NOCELLA, G. Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web, **Meta**, 2004, disponível em

<<https://www.erudit.org/revue/meta/2004/v49/n2/009351ar.pdf>> Acesso em 9 jul. 2015.

GILE, D. Testing the Effort Models' tightrope hypothesis in simultaneous interpreting – A contribution. **Hermes, Journal of Linguistics**, n. 23, 1999, disponível em: <<https://pt.scribd.com/document/289097147/Tightrope-Hypothesis-Hermes>> Acesso em: 7 nov. 2017.

HOLUB, E. Does Intonation Matter? The impact of monotony on listener comprehension. **The Interpreters' Newsletter**, n. 15, p. 117-126, 2010.

KURZ, I. **Conference Interpreting – User Expectations, Coming of Age**. Proceedings of the 30th Annual Conference of the American Translators Association (D. L. Hammond, ed.), Medford (NJ), Learned Information, 1989.

LENGLET, C. The impact of prosody on the comprehensibility and quality of simultaneous interpreting: A pilot study. **Translation and Comprehensibility**, 2015, disponível em <<http://di.umons.ac.be/details.aspx?pub=74a252b8-ed5-4c1e-93d9-af4c2c9f0974>> Acesso em: 9 julho 2015.

MOSE, P. Expectations of Users of Conference Interpretation. **Interpreting**, 1(2), p. 145-178, 1996.

PÖCHHACKER, F. & ZWISCHENBERGER, C. **Survey on quality and role: conference interpreters' expectations and self-perceptions**, 2010, disponível em <<http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1>> Acesso em: 9 julho 2015.

RENNERT, S. The impact of fluency on the subjective assessment of interpreting quality. **The Interpreters' Newsletter**, n. 15, p. 101-115, 2010.

SCHLESINGER, M. Intonation in the production of and perception of simultaneous interpretation. In: **Bridging the Gap. Empirical Research in Simultaneous Interpretation**. S. Lambert; B. Moser-Mercer (Eds.) Benjamins, 1994.

TAUROZA, S. & ALLISON, D. Speech Rates in British English. **Applied Linguistics**, 11 (1): p.90-105, 1990.

VANDER ELST, P. **The Nuremberg Trial**. 2000, disponível em: <<http://aiic.ca/page/983/the-nuremberg-trial/lang/1>> Acesso em: 9 julho 2015.

ZACKIEWICS, D.V.; ANDRADE, C.F. Seis parâmetros de fluência. **Revista da Sociedade Brasileira de Fonoaudiologia**: 5(7), p. 59-64, 2000.

Resumo

O objetivo deste estudo foi verificar as características prosódicas da fala semiespontânea e interpretada e analisar como as características da fala interpretada podem influenciar a avaliação de qualidade e compreensão da mensagem produzida. As falas foram analisadas acusticamente em termos de número e distribuição de grupos entoacionais e pausas, taxa de elocução e frequência fundamental. Noventa juízes avaliaram os efeitos impressionantes da produção das intérpretes e três formadores de intérpretes compararam o discurso interpretado com o de partida. Os resultados foram discutidos para verificação de similaridades e diferenças entre as avaliações.

Palavras-chave: Interpretação simultânea; prosódia; entonação; qualidade; percepção da fala.

Abstract

The purpose of this study was to verify the prosodic features of semi-spontaneous and interpreted speeches and analyze how such features may interfere in the assessment of quality and comprehension of the message delivered. Deliveries were analyzed in terms of number and distribution of intonation groups and pauses; speech rate; and fundamental frequency rates. Ninety judges assessed the impressive effects of interpreters' oral production. Three interpreter trainers compared the interpreted speech to the original lecture. Results were then compared for similarities and differences.

Keywords: Simultaneous interpreting; prosody; intonation; quality; perception of speech.