



SOLICITANTE: ANTONIO MARIZ DE OLIVEIRA, ADVOGADO

I) MATERIAL QUESTIONADO

Ao perito signatário foi apresentado um arquivo de áudio intitulado "PR1 14032017", cujas características, verificadas por meio do programa *MediaInfo* encontram-se abaixo listadas.

General	
Complete name	: C:\PR1 14032017.WAU
Format	: Wave
File size	: 35.7 MiB
Duration	: 38mn 56s
Overall bit rate mode	: Constant
Overall bit rate	: 128 Kbps
Audio	
ID	: 0
Format	: ADPCM
Format settings, Endianness	: Little
Codec ID	: 11
Codec ID/Hint	: Intel
Duration	: 38mn 56s
Bit rate mode	: Constant
Bit rate	: 128 Kbps
Channel(s)	: 2 channels
Sampling rate	: 16.0 KHz
Bit depth	: 4 bits
Stream size	: 35.7 MiB (100%)

II) OBJETIVOS PERICIAIS

Ao perito signatário foram apresentados quesitos específicos, os quais serão respondidos na seção V.

Considerando a complexidade do caso em tela, entendeu o perito signatário ser necessária uma discussão prévia, a qual será desenvolvida na seção IV.



III) INSTRUMENTAL UTILIZADO

- programa *PRAAT*
- programa *MultiSpeech*
- programa SPG 16.0
- programa *SoundForge*

IV) DISCUSSÃO

Em todo caso de grande repercussão e de potenciais desdobramentos políticos perde-se facilmente a objetividade e as paixões costumam falar mais alto do que a razão. Não raramente, procedimentos e protocolos técnicos consagrados e universalmente aceitos são distorcidos ou interpretados de forma "flexível", de modo a atender interesses muito pouco nobres.

No caso em tela temos um exemplo típico dessa situação. Uma gravação inteiramente contaminada por inúmeras descontinuidades, mascaramentos por ruído, longos trechos ininteligíveis ou de inteligibilidade duvidosa e várias outras incertezas não poderia ser considerada como uma prova material válida.

Não existe prova "mais ou menos" boa. **A prova, tecnicamente, é boa ou não é. E a gravação questionada, no estrito sentido técnico-pericial, deveria ter sido considerada imprestável desde o primeiro momento.** Nenhum perito judicial, no bojo de um processo corriqueiro, consideraria esta gravação como válida, pelo simples fato de ser impossível garantir sua integridade e de ser impossível garantir que não houve trechos extirpados por edição *a posteriori*, de ser inclusive impossível garantir que alguns sons não pudessem ter sido inseridos em pós processamento digital.



Ressalte-se que não falamos de um ponto isolado, de uma descontinuidade cuja raridade permitiria, com alguma liberalidade, que fosse desprezada. Não, pelo contrário, falamos aqui de uma gravação eivada de cortes, ataques abruptos, ruídos interferentes, largos trechos ininteligíveis etc., vícios que foram, sem auxílio de qualquer aparelhagem sofisticada, perceptualmente detectadas por milhares de pessoas.

A pergunta que se considera importante é porque a PGR se apressou em publicizar uma gravação tão nitidamente corrompida antes de submetê-la a uma perícia técnica rigorosa. Com efeito, a própria ASSOCIAÇÃO NACIONAL DOS PERITOS CRIMINAIS manifestou-se veementemente a esse respeito:

"A Associação Nacional dos Peritos Criminais Federais esclarece que, ao se ouvir o áudio divulgado pela imprensa, percebe-se a presença de eventos acústicos que precisam passar por análise técnica, especializada e aprofundada, sem a qual não é possível emitir qualquer conclusão acerca da autenticidade da gravação. Ademais, sempre que houver vestígios materiais, é temerária a homologação de delações sem a devida análise pericial

[<http://apcf.org.br/Noticias/AgenciaAPCF/tabid/341/post/nota--per-cia-da-pf-recomenda-an-lise-de-udio-que-envolve-o-planalto/Default.aspx>]

Não vai aqui qualquer concordância prévia com eventuais conclusões a serem emitidas no futuro pelo INC. Mas não há como negar que a observação da APCF é pertinente. O fato é que a gravação, apesar de seus vícios gritantes, foi divulgada com inusitado açodamento, tendo como (pseudo) fundamento pericial uma simples impressão subjetiva de não especialistas, os quais emitiram a Informação 030/2017, em 07/004/2017 (ver figura 01), na qual apresentam uma descrição confusa e reconhecidamente superficial (visto que "meramente perfunctória", como afirmado explicitamente na referida informação), referindo-se vagamente a "características iniciais de confiabilidade" (eufemismo bastante inadequado para substituir "autenticidade tecnicamente comprovada").



O mais grave, no entanto, na referida Informação do MPF, não é o uso de terminologia ambígua, mas antes conter afirmações tecnicamente insustentáveis e que, na verdade, conflitam frontalmente com a realidade revelada pela gravação. É, para dizer pouco, imprudente e equivocada a conclusão de que a gravação "encontra-se audível, apresentando sequência lógica". Há tantos equívocos nessa afirmação que é preciso analisá-la por partes. Em primeiro lugar, confundiu-se os conceitos de *audibilidade* e *inteligibilidade*.

A figura 02 ilustra esse aspecto. A razão sinal/ruído em diversos pontos da gravação, em especial em trechos com a fala do presidente Michel Temer, o qual se encontrava afastado, no mínimo, alguns metros do gravador, é muito baixa. No espectrograma mostrado na figura 02 é analisado o trecho a seguir (um dos trechos "polêmicos"):

JOESLEY: tô no meio [e não "todo mês", como equivocadamente divulgado pela PGR] - setas amarelas

PRESIDENTE TEMER: ininteligível - setas verdes

JOESLEY: também... - setas vermelhas

Fica claro, na análise espectrográfica, que a intervenção do presidente Michel Temer tem uma razão sinal/ruído extremamente desfavorável. Praticamente inexistente informação acima de 500 Hz (região abaixo da linha vermelha horizontal). Observe-se que há uma diferença acústica muito grande entre a intervenção do presidente Temer e as falas de Joesley, sendo que estas apresentam harmônicos ainda com alguma definição na faixa de 2.5 kHz.

Apesar de se saber que tal trecho (entre as setas verdes) contém fala do presidente, o trecho é totalmente ininteligível, nada havendo em termos de filtragem que pudesse sanar essa deficiência, visto que, supostamente, se trata de um problema na captação do áudio, durante a qual, seja pela distância e/ou direcionamento da fonte sonora (posição relativa microfone/locutor) ou mesmo características intrínsecas do aparelho, o trecho se tornou



ininteligível. Em muitos casos irremediavelmente ininteligível. O trecho é, entretanto "audível", qualidade que, na prática, pouca importância tem.

Não foi possível aferir exatamente qual a razão sinal ruído exata no trecho comentado, mas certamente temos uma razão negativa, ou seja, o nível do ruído suplanta em muito o nível da fala, como fica evidente no espectrograma da figura 02. É interessante consultar o estudo de Miller *et. al.* (1951¹), o qual continua atual. O gráfico abaixo demonstra a queda muito significativa de inteligibilidade de palavras em frases. De zero (ou seja, nível do ruído igual ao do sinal de fala) para -12 dB a inteligibilidade cai de cerca de 60% para praticamente zero.

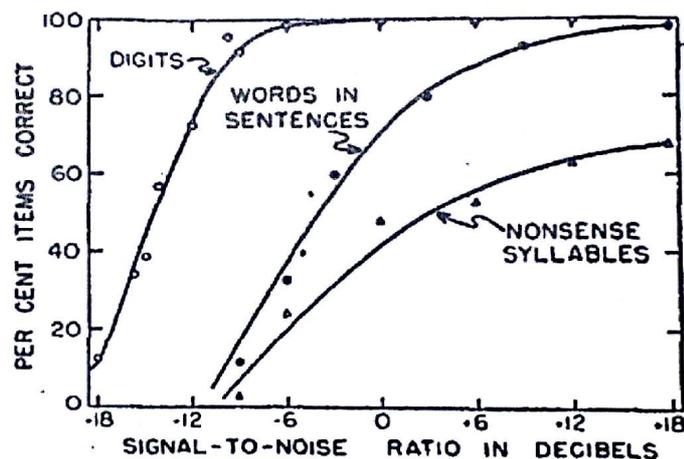


FIG. 1. Relative intelligibility of different test materials

No caso em tela uma razão sinal/ruído de -12dB seria até otimista, no trecho analisado na figura 02. Não causa qualquer surpresa, portanto, que tal trecho, assim como inúmeros outros com as mesmas características ao longo da gravação, seja

¹ The intelligibility of speech as a function of the context of the test materials. Miller, George A.; Heise, George A.; Lichten, William, *Journal of Experimental Psychology*, Vol 41(5), May 1951, 329-335



ininteligível em toda a sua extensão e que assim permaneça - ininteligível - independentemente de qualquer tipo de filtragem que se queira aplicar.

Como fica claro no espectrograma da figura 02, o ruído interferente se estende ao longo de uma larga faixa espectral, o que impede a aplicação de filtros do tipo *noise reduction* ou *noise gate*, pois se tal procedimento for eventualmente empregado, não haverá qualquer ganho de inteligibilidade.

A afirmação, portanto, contida na Informação 030/2017 do MPF, de que algumas falas "tomam-se incompreensíveis sem a utilização de equipamentos especializados" remete a um otimismo bastante ingênuo e que não se ajusta à realidade do áudio questionado. Alguns trechos são "incompreensíveis" e assim permanecerão.

É importante, pois, se fazer a distinção entre os conceitos de *audibilidade* e *inteligibilidade*, para não incorreremos no mesmo erro dos técnicos do MPF. Em outras palavras: uma fala pode ser *audível* (inclusive sendo possível saber quem a produziu) mas ao mesmo tempo ser *ininteligível*. Tal distinção, entretanto, não foi devidamente esclarecida na Informação 030/2017 do MPF, na qual se afirma que a gravação "encontra-se audível, apresentando sequência lógica".

Ora, como vimos, ser "audível" não é sinônimo de ser "inteligível". Aparentemente a "analista" e o "técnico" que emitiram a referida Informação não conhecem a distinção entre os dois conceitos. E mais: causa uma enorme estranheza que os emitentes da referida informação possam afirmar tão categoricamente que a gravação apresenta "sequência lógica" se existem tantos trechos ininteligíveis, ruídos interferentes etc. Em especial nas falas do presidente Michel Temer, em função de seu maior afastamento em relação ao gravador e de eventuais mudanças de direcionalidade em relação ao microfone captador, a razão sinal/ruído frequentemente se torna crítica, destruindo totalmente a inteligibilidade. Nessa situação é preciso cautela para usar critérios vagos como "sequência lógica", certamente muito dependentes de avaliações subjetivas. Por



óbvio não é possível garantir o encadeamento discursivo lógico se um dos interlocutores tem a maior parte de suas falas ininteligível. Além disso vale notar que em conversações coloquiais, pode haver mudanças rápidas de tópico conversacional, o que permite que se extraiam trechos originais sem que o efeito final, no domínio daquela interação, possa resultar anômalo quanto à uma impressão subjetiva em termos de coesão e coerência.

Mas a baixa inteligibilidade não é o único nem o mais grave problema que macula a gravação questionada. **Existem inúmeros pontos ao longo da gravação nos quais se poderia efetuar, sem deixar qualquer vestígio, uma edição envolvendo corte de material original.** Trata-se de descontinuidades com perda de material fonético, regiões com mascaramento total do sinal de interesse por ruído sobreposto e regiões com *clipping*. Em qualquer uma dessas situações seria possível criar um ponto de edição, retirar material original e finalmente recompor a forma de onda, aproveitando a região do ruído, *clipping* ou perda de material fonético para encobrir o ato fraudulento.

Não tivemos oportunidade de examinar diretamente o gravador que realizou a captação dessa conversação descontinuidade. Pode ser que alguns dos efeitos de edição tenham sido gerados pelo gravador, aparentemente um equipamento de muito baixa qualidade (tanto assim que teria registrado o sinal com uma profundidade de 4 bits, algo raro de se encontrar, mesmo em equipamentos de baixo custo). O que é importante frisar é que o fato de o gravador possuir alguma falha sistêmica eventualmente associada aos eventos de descontinuidade não justificaria todos os pontos suspeitos que encontramos.

Na verdade, **as descontinuidades não ocorrem de modo previsível.** Não há, por exemplo, uma associação direta entre a amplitude do sinal e a presença da descontinuidade. Os eventos ocorrem tanto em plosivas quanto em vogais, e nos dois casos de modo não previsível. Há pontos com plosivas de grande intensidade e que, ainda assim, não criaram descontinuidades. A figura 06 ilustra este aspecto.



No caso das vogais ocorre o mesmo. Algumas vogais aparecem truncadas enquanto outras, com a mesma intensidade no ataque revelam um aumento gradual de intensidade, tal como esperado em qualquer gravação (ver figura 05).

O comportamento caótico e sem um padrão discernível no aparecimento das discontinuidades aumenta a suspeição sobre a gravação questionada. Outro aspecto a considerar é a não homogeneidade na distribuição dos eventos de discontinuidade. Apenas entre os momentos 11:36 e 11:53, cerca de 17 segundos, por exemplo, foram detectados cinco pontos de possível edição, sendo essa região a de maior concentração desse tipo de evento acústico. Curiosamente, o mesmo trecho contém as falas mais divulgadas e exploradas pela mídia.

Assim, temos uma série de características, na gravação questionada, que não permitem que se garanta sua autenticidade: **(a) discontinuidades não justificadas; (b) ruídos muito intensos mascarando totalmente o sinal de interesse; (c) clipping (saturação) e (d) longos trechos ininteligíveis.**

A prova é tecnicamente ruim. A prova é imprestável, visto que não se pode garantir sua autenticidade. O recurso a horários de programação de rádio como base para garantir autenticidade é inaceitável. Uma simples edição ao final da gravação permitiria tal ajuste. Com efeito, aos 35:25,663, perto do final da gravação e antes da entrada da programação da CBN, ocorre uma discontinuidade com vários pontos de *clipping*, região na qual se poderia efetuar um corte indetectável de modo a "ajustar" a gravação ao horário da rádio (ver figura 07). Como não há contexto conversacional no trecho, a ação poderia ser realizada sem qualquer dificuldade.

A presença de muitos pontos com saturação permitiria também que, caso se retirasse material original, o tempo total pudesse ser ajustado aumentando artificialmente a duração de regiões de ruído. Outro artifício que levaria ao mesmo resultado seria o alongamento artificial de trechos constituídos exclusivamente de ruído de fundo de baixa



amplitude, um procedimento simples e também indetectável (ressalte-se que há muitos trechos com tais características ao longo da gravação).

As figuras 08 e 09 reúnem diversos detalhes de oscilograma em pontos nos quais uma edição pode ter sido praticada. As duas figuras estão longe de ser exaustivas, visto haver muitos outros pontos com as mesmas características.

V) RESPOSTAS AOS QUESITOS

01) *Queira o perito descrever o material questionado.*

Resposta: Ver item I.

02) *A gravação apresentada pelo delator é original?*

Resposta: Não. De acordo com a Informação 0230/2017 do MPF, o delator Joesley entregou um *pendrive* com quatro arquivos de áudio intitulados "AEunique.WAV", "PR1 14032017.WAV", "PR2 16032017.WAV" e "PR2 A 13032017.WAV". Reproduzimos abaixo o trecho da Informação 030/2017 no qual são descritos os arquivos, tal como entregues ao MPF.

Esta Secretaria recebeu para exame uma mídia removível *pen drive* contendo uma (1) pasta e 4 (quatro) arquivos, conforme figura apresentado no *Windows Explorer*.

- 1) AEunique.WAV
- 2) PR1 14032017.WAV
- 3) PR2 16032017.WAV
- 4) PR2 A 13032017.WAV

Os arquivos entregues não podem ser os originais, visto que, por óbvio, tiveram seus nomes alterados. Nenhum dispositivo gravador gera nomes nesses formatos. Os nomes devem ter sido atribuídos pelo próprio delator. É estranho que o MPF não tenha tido o cuidado de solicitar os arquivos originais.



03) Há indícios de edição na gravação?

Resposta: Existem dezenas de pontos de descontinuidade ao longo de toda a gravação. Cada um desses pontos é, potencialmente, um ponto de edição. Sabe-se que nem todas as descontinuidades existentes na gravação são pontos de edição, visto que são muito numerosas. É possível, pois, que algumas dessas descontinuidades seja decorrente de alguma falha sistêmica do aparelho gravador. A questão pericialmente relevante, no entanto, não é se o aparelho gravador eventualmente produz tais falhas, **mas sim que tais falhas abrem a porta para que sejam realizadas edições cuja detecção seria impossível, visto que uma edição feita com algum cuidado poderia reproduzir exatamente as características de uma falha sistêmica real.** Vale notar que critérios baseados em continuidade discursiva não são aplicáveis em muitos pontos onde ocorrem descontinuidades, até porque grande parte das falas do presidente Michel Temer é ininteligível. Seria temerário valer-se de um critério baseado em "sequência lógica", como fez o MPF, em uma conversação, na qual, na maior parte do tempo, entende-se apenas a voz de um dos interlocutores (Joesley).

Ressalte-se que não são apenas as descontinuidades evidenciadas por perdas fonéticas que poderiam ser considerados potenciais pontos de edição. A gravação está repleta de ruídos muito intensos que praticamente ocupam toda a faixa do espectro, mascarando completamente o sinal de interesse. Existem também diversos pontos de *clipping* (saturação), também candidatos a ocultar cortes fraudulentos de forma indetectável.

04) A gravação pode ser considerada autêntica?

Resposta: Não. Do ponto de vista técnico-pericial, a gravação apresenta demasiados vícios para que possa ser considerada autêntica. Não faz qualquer sentido acatar como autêntica uma gravação que contém dezenas de descontinuidades, pontos de clipping, ruídos de mascaramento e longos trechos ininteligíveis. Em qualquer processo corriqueiro



esse tipo de gravação é descartada por possuir demasiados indícios de possível edição. Não cabe ao perito especular, admitindo hipóteses pouco ortodoxas: **uma descontinuidade é sempre um potencial ponto de edição e nada pode mudar esse fato, ainda que se descubra que falhas sistêmicas possam provocar eventuais descontinuidades.** Vivemos um mundo digital e manipulações envolvendo trechos ruidosos do sinal são bastante simples com os recursos de processamento digital hoje disponíveis. Assim, qualquer indício de edição deve ser assim tratado, salvo existir prova sólida que contradiga tal hipótese. No caso em tela, algumas descontinuidades podem ser descartadas em função da continuidade discursiva, mas restam muitas para as quais tal critério não seria aplicável.

05) A gravação, do ponto de vista jurídico, poderia ser usada como prova?

Resposta: Não. Como já esclarecido no quesito anterior, não se pode, por nenhum meio, garantir a autenticidade da gravação, sendo, portanto, a prova imprestável para fins judiciais. Para além dos indícios muito consistentes de edição vale lembrar que grande parte da gravação, mais especialmente no que diz respeito às falas do presidente Temer, é ininteligível, dificultando, quando não impossibilitando, a definição inequívoca dos contextos em que cada fragmento inteligível ocorre.

Goiânia, 22 de maio de 2017


Prof. Dr. Ricardo Molina de Figueiredo


Maurício Raymundo de Cunto
CREA. 060.154.048-7

**MPF****Ministério Público Federal****Procuradoria-Geral da República
Gabinete do Procurador-Geral da República
Secretaria de Pesquisa e Análise – SPEA/PGR****SIGILOSO**

Informação nº 030/2017-SPEA/PGR

07 de abril de 2017.

Para a verificação do arquivo, utilizou-se a ferramenta *Audacity*, software de áudio livre e código aberto com suporte ao formato de áudio .wav.

2 ANÁLISE DE COERÊNCIA PERCEPTIVA

O arquivo apresentado foi analisado de forma preliminar, submetido a oitiva sob a perspectiva exclusiva da percepção humana. Não houve auxílio de equipamentos especializados na avaliação sobre a integridade dos áudios.

Nesse contexto, o objetivo do trabalho foi verificar se os diálogos existentes nos áudios estão inteligíveis e, se numa análise meramente perfunctória, os arquivos possuem ou não características iniciais de confiabilidade.

O diálogo constante no Áudio 02 (PR1 14032017.WAV) encontra-se audível, apresentando sequência lógica. O arquivo possui alguns ruídos e a voz de um dos interlocutores apresenta-se com maior intensidade em relação à voz do segundo interlocutor, e em alguns momentos, tornam-se incompreensíveis sem a utilização de equipamentos especializados.

Figura 01

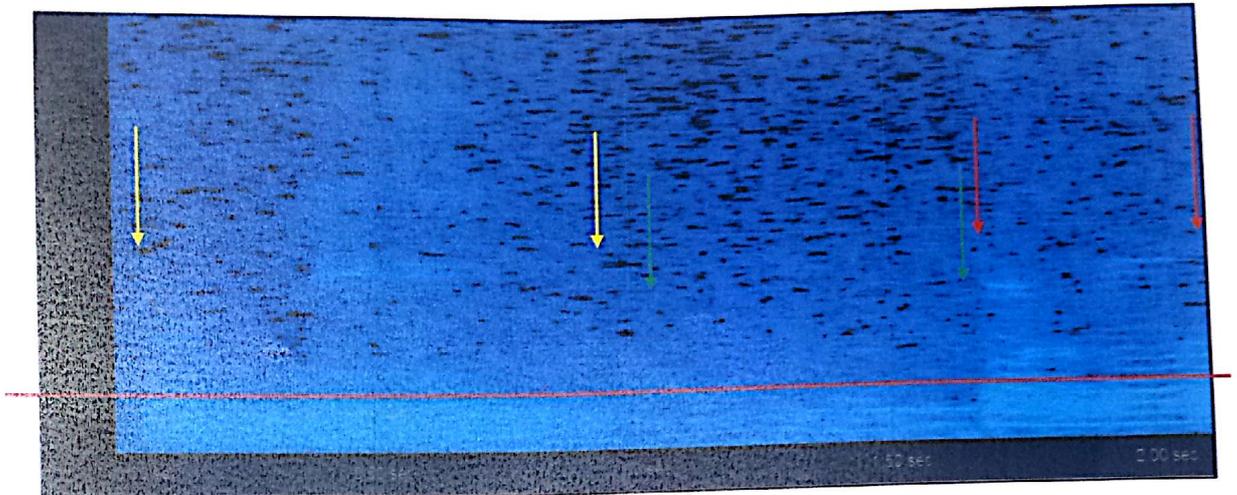


Figura 02. A razão sinal/ruído em diversos pontos da gravação, em especial em trechos com a fala do presidente Temer, o qual se encontrava afastado, no mínimo, alguns metros do gravador, é muito baixa. No espectrograma mostrado na figura acima é analisado o trecho a seguir:

JOESLEY: tô no meio [e não "todo mês", como equivocadamente divulgado pela PGR] - setas amarelas

PRESIDENTE TEMER: ininteligível - setas verdes

JOESLEY: também... - setas vermelhas

Fica claro que a intervenção do presidente Temer tem uma razão sinal/ruído extremamente desfavorável. Praticamente inexistente informação acima de 500 Hz (região abaixo da linha vermelha horizontal). Observe-se que há uma diferença acústica muito grande entre a intervenção do presidente Temer e as falas de Joesley, sendo que estas apresentam harmônicos ainda com alguma definição na faixa de 2.5 kHz.

Apesar de se saber que tal trecho (entre as setas verdes) contém fala do presidente, o trecho é ininteligível, nada havendo, em termos de filtragem que pudesse sanar essa deficiência, visto que, supostamente, se trata de um problema na captação do áudio, durante a qual, seja pela distância e/ou direcionamento da fonte sonora (posição relativa microfone/locutor) ou mesmo características intrínsecas do aparelho, o trecho se tornou ininteligível. Em muitos casos irremediavelmente ininteligível.

É importante, pois, se fazer a distinção entre os conceitos de *audibilidade* e *inteligibilidade*. Em outras palavras: uma fala pode ser *audível* (inclusive sendo possível saber quem a produziu) mas ao mesmo tempo ser *ininteligível*. Tal distinção, entretanto, não foi devidamente esclarecida na Informação 030/2017 do MPF, na qual se afirma que a gravação "encontra-se audível, apresentando sequência lógica".

Ora, como vimos, ser "audível" não é sinônimo de ser "inteligível". Aparentemente a "analista" e o "técnico" que emitiram a referida Informação não conhecem a distinção entre os dois conceitos. E mais: causa uma certa estranheza que os emitentes da referida Informação possam afirmar tão categoricamente que a gravação apresenta "sequência lógica" se existem tantos trechos ininteligíveis, ruídos interferentes etc.

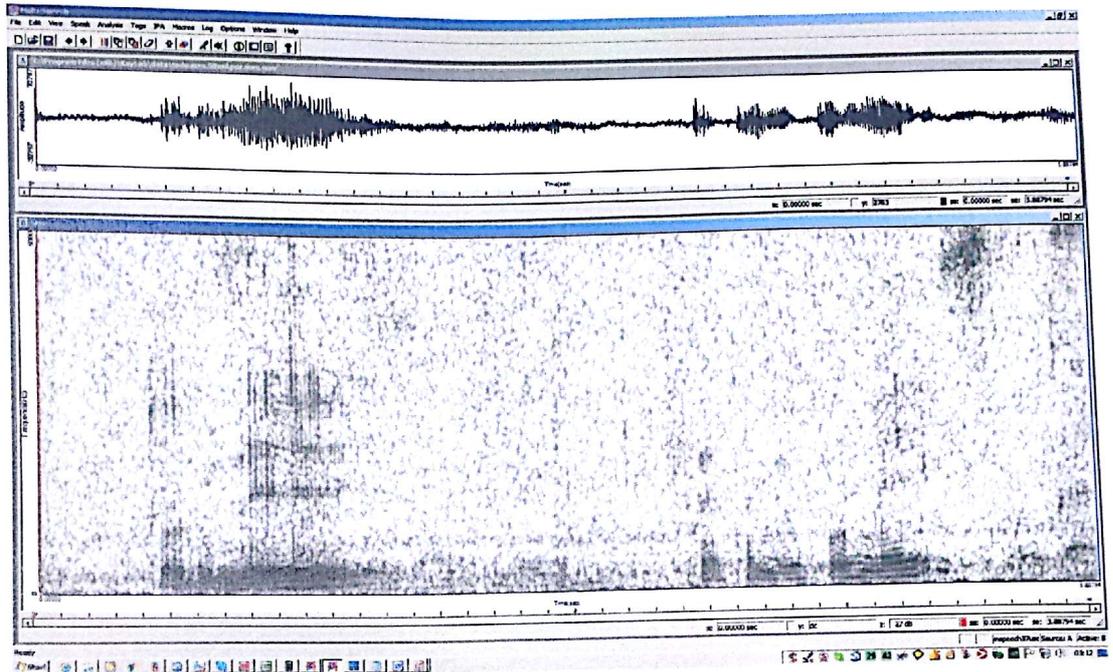


Figura 03. "tô no meio" vs. "50 mil por mês" (momentos 11:41, 296 e 15:38,612, respectivamente). Vê-se claramente que a expressão produzida no [primeiro caso foi "tô no meio" e não "todo mês", visto que inexistente vestígio da fricativa ao final, enquanto esta aparece claramente no segundo caso, mesmo que a amplitude em "mês" tenha sido menor.

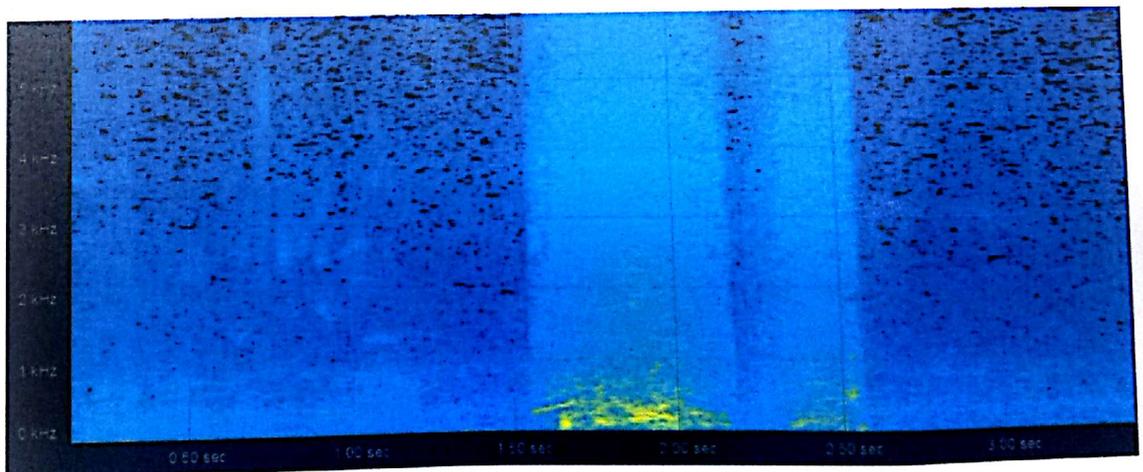


Figura 04. Em diversos pontos há a ocorrência de ruídos de tal intensidade, que chegam a cobrir praticamente toda a faixa do espectro, tal como ocorre no trecho analisado no espectrograma acima (aos 10:16,381).

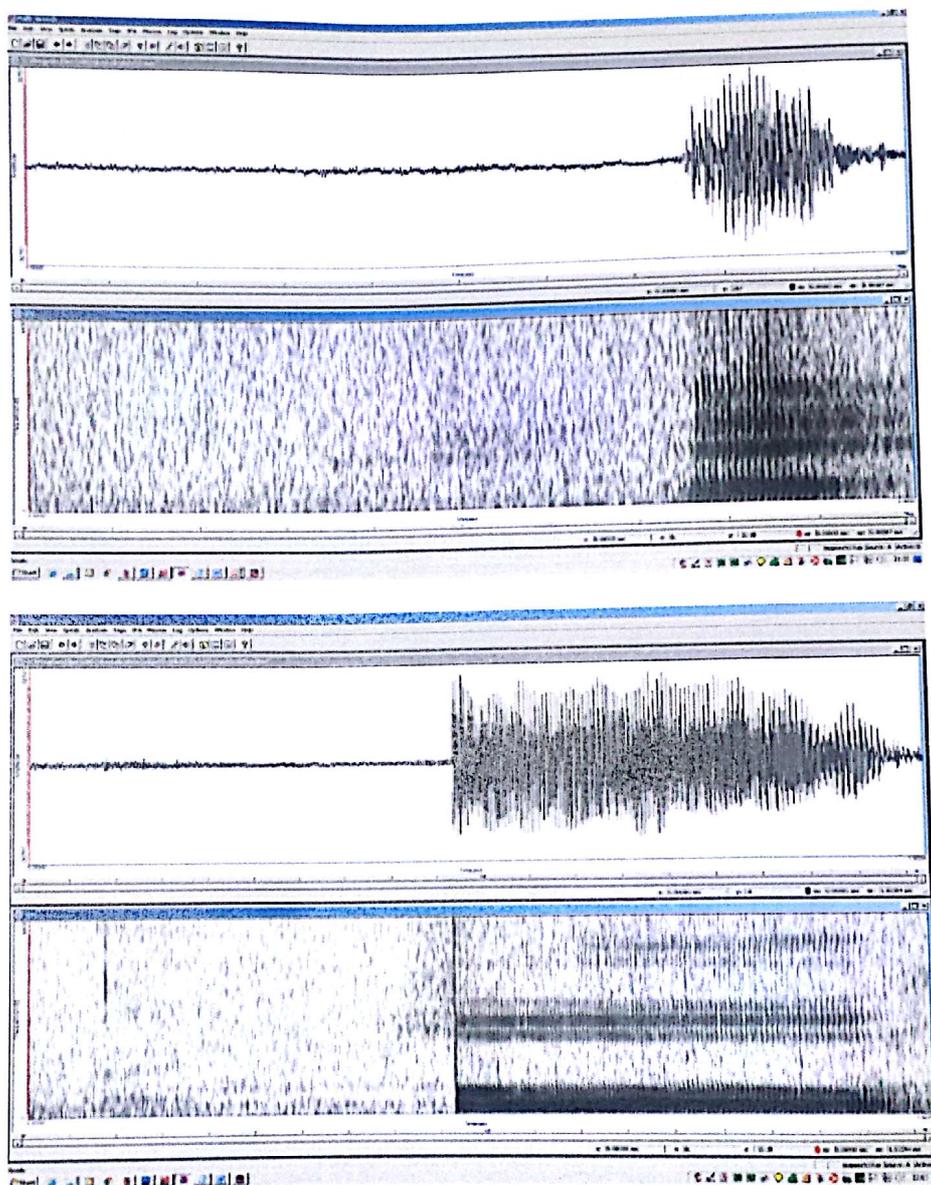


Figura 05. Os eventos de descontinuidade não ocorrem de forma regular nem são previsíveis em função do contexto fonético. Nos espectrogramas acima analisamos duas vogais de alta amplitude. No quadro superior trata-se da vogal /E/ (aos 07:30,252) e no inferior a vogal /o/ (aos 07:13,527). As duas atingem picos semelhantes de amplitude, sendo que o /E/ tem o ataque normal e a vogal /o/ revela ataque truncado com perda de material fonético original. A eventualidade de ser tal comportamento, em alguns casos, devido a eventual efeito sistêmico (algo a se conferir, é claro), de modo algum transformaria a gravação em uma prova autêntica e juridicamente aceitável. Qualquer descontinuidade é um virtual ponto de edição.

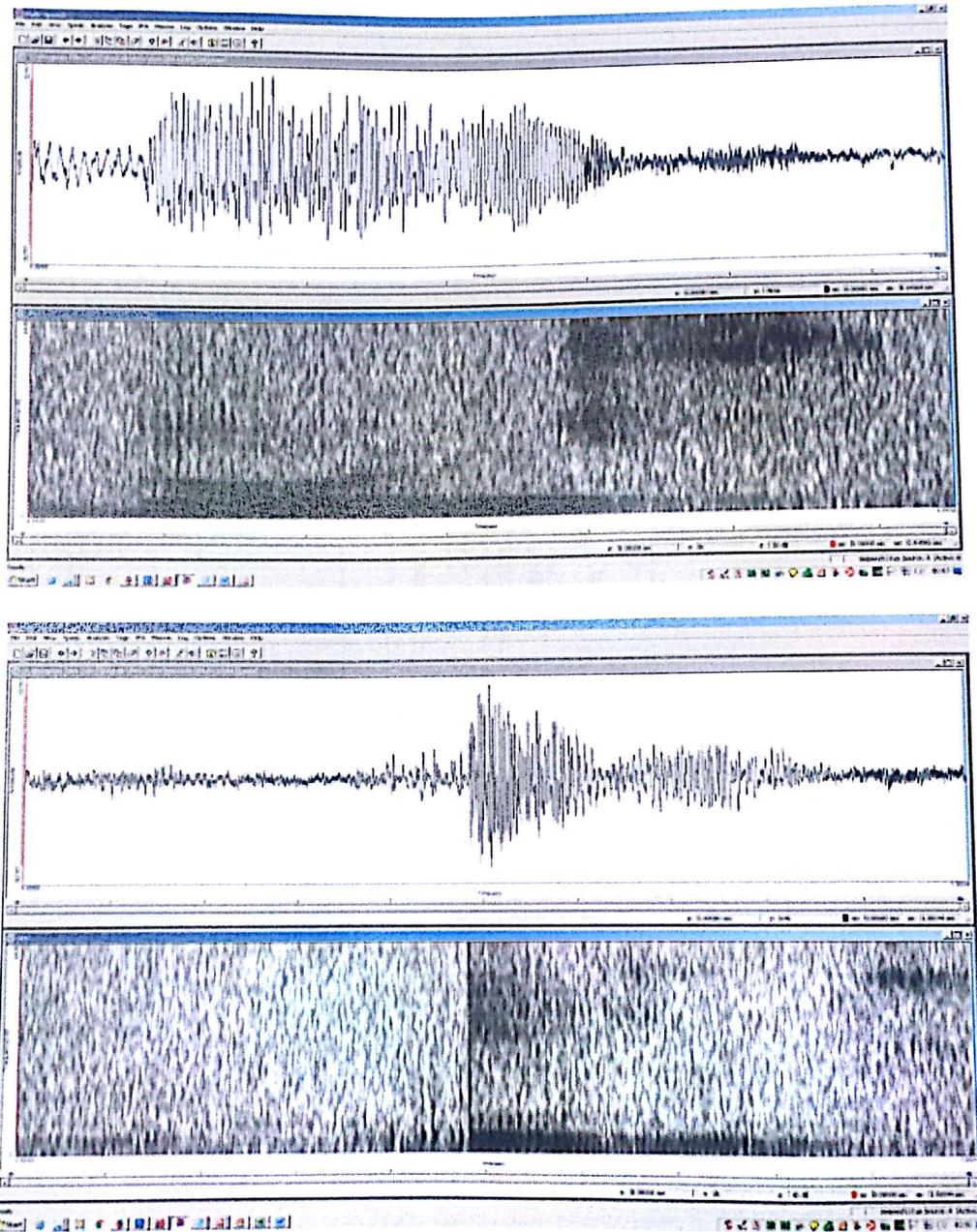


Figura 06. Assim como ocorre com as vogais, as discontinuidades associadas a perda de material fonético envolvendo plosivas também não apresenta um padrão regular. No espectrograma superior vemos a análise da palavra "dois" (aos 12:09,330), sem criação de descontinuidade e no inferior a da palavra "doze"(aos 07:27,615), com descontinuidade.

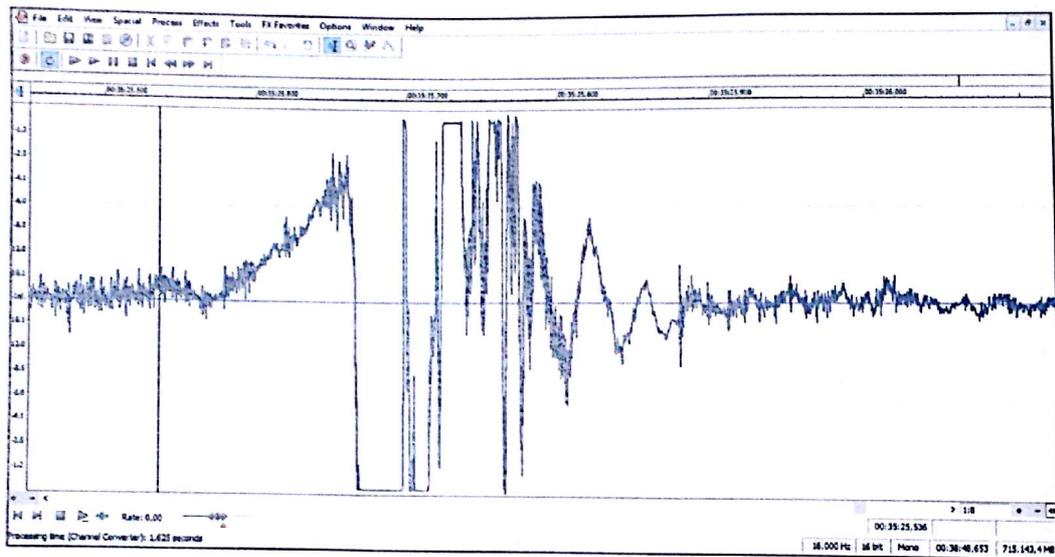


Figura 07. Região com clipping próxima ao final da gravação (em 35:35,663), pouco antes da entrada da gravação da CBN. Este é um potencial ponto de edição

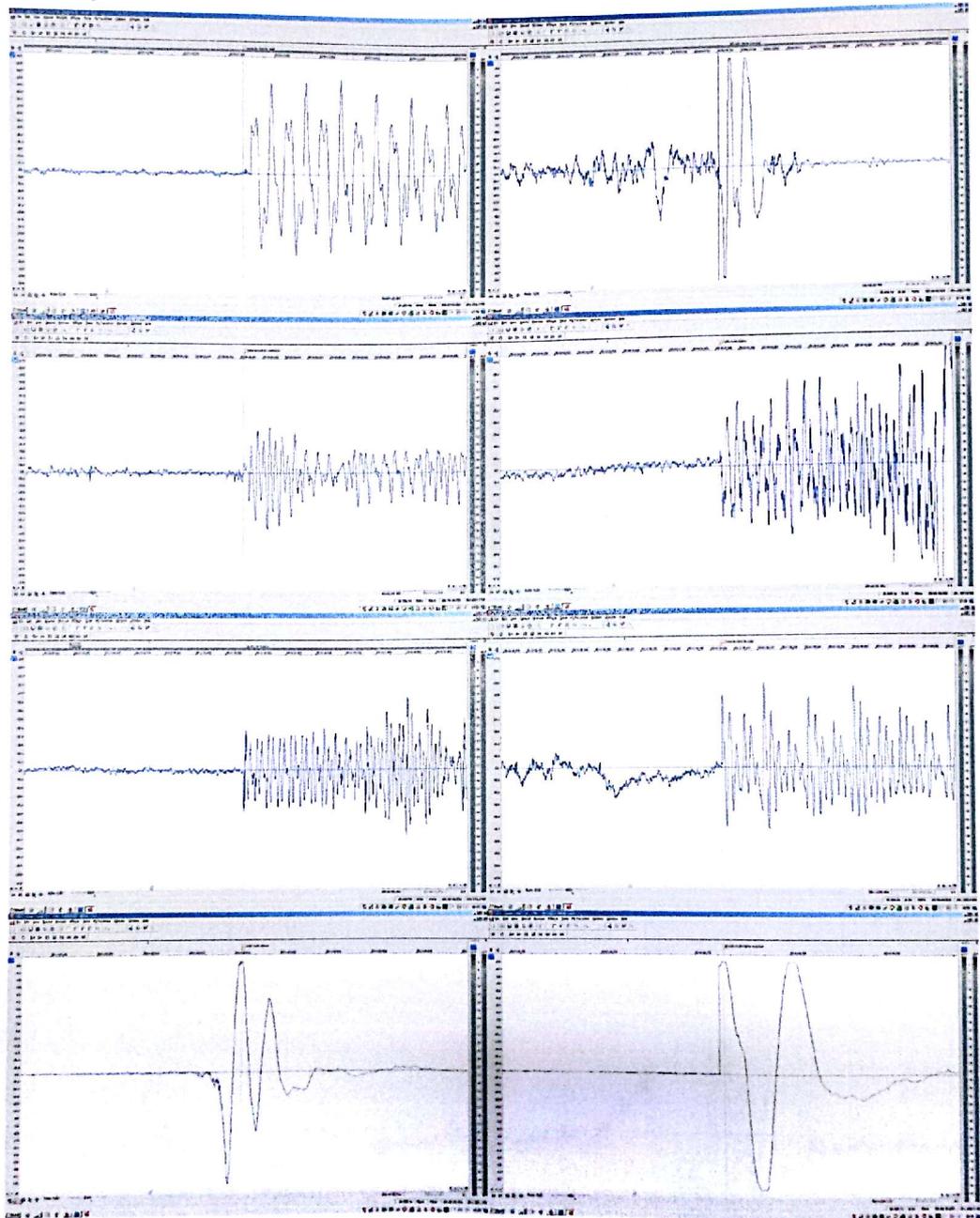


Figura 08.

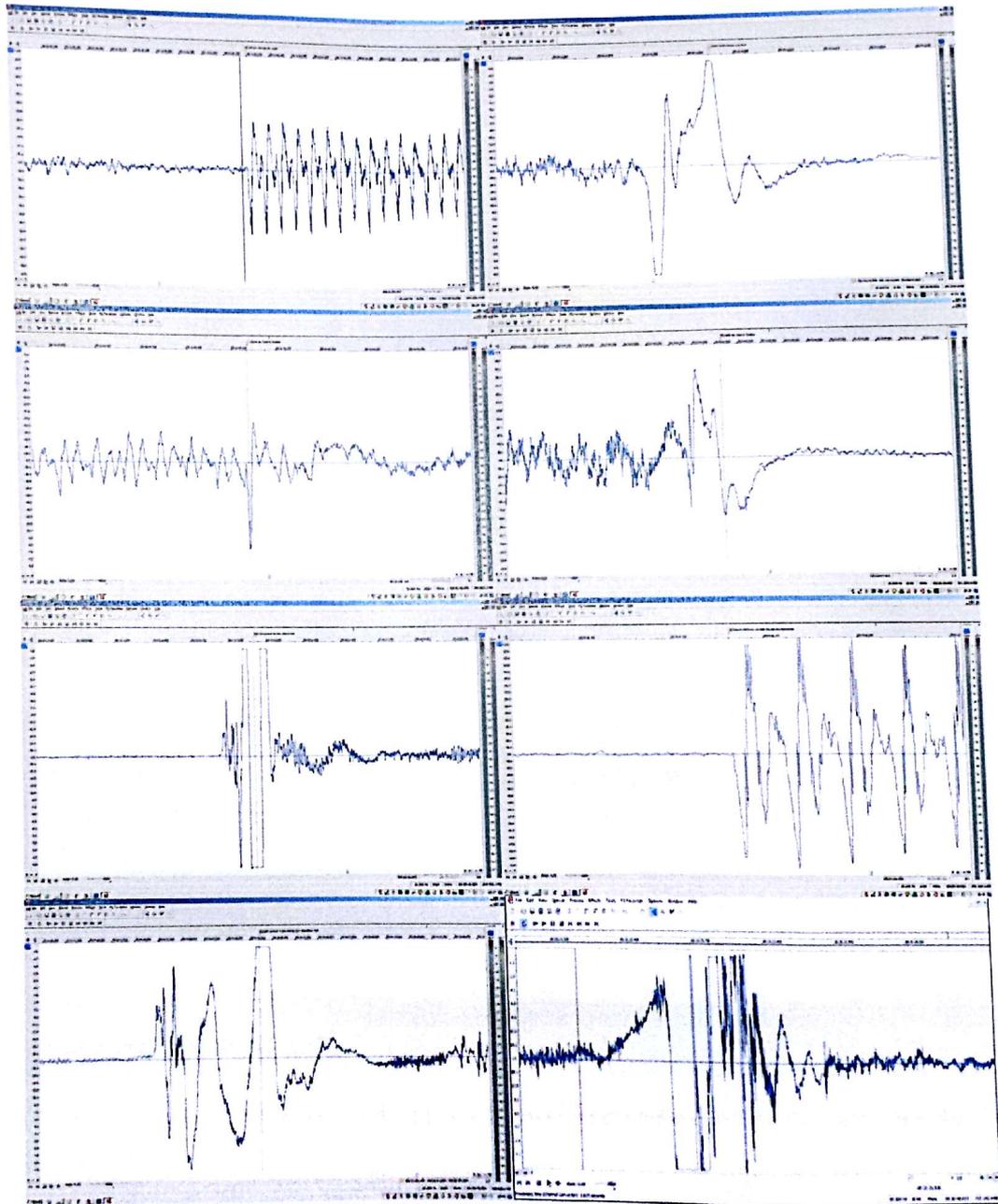


Figura 09