
MOBILE TRUST
TELECOMMUNICATIONS

Índice

Prefácio

Este manual apresenta e comprova a confiabilidade das conversões de voz com o intuito de encriptação desenvolvidas para serem usadas no canal vocal das redes de telefonia móvel e Skype.

1. Termos e definições

Esta seção contém termos e definições básicos utilizados nas seções porteriores. Outros termos serão explicados no próprio texto.

Trechos do sinal de voz – uns sons de voz selecionados aleatoriamente ou algumas partes dos mesmos limitadas por tempo e freqüência. É possível manter esses trechos inteiros dentro do sinal encriptado, no entanto, eles devem ser bastante curtos para impedir a identificação do significado da voz.

Segurança criptográfica das conversões de encriptação - alta improbabilidade a uma possível recuperação de dados encriptados por um intruso. Com a alta segurança criptográfica a recuperação do discurso original é impossível ou extremamente difícil.

Desencriptação - recuperação das informações encriptadas. Quando a desencriptação é realizada sem a chave verdadeira, o objetivo do intruso pode ser a recuperação do significado do discurso enquanto o próprio sinal de voz original pode ser restaurado com distorções.

Inteligibilidade residual da voz - a percentagem das unidades de voz corretamente identificáveis durante a escuta do sinal encriptado. Na maioria das vezes as unidades de voz são palavras, frases ou sílabas. Se a segurança das conversões de encriptação for alta a inteligibilidade residual é próxima de zero.

Camuflagem de voz – um conceito utilizado em conjunto com o conceito de inteligibilidade residual, mas tem sentido oposto, ou seja, quanto menor a inteligibilidade residual maior a camuflagem de voz.

Volume do trecho do sinal de voz – o resultado da multiplicação do comprimento do

trecho do sinal de voz pela largura da faixa de freqüências dele. Este valor é adimensional e caracteriza a complexidade das conversões de encriptação. Quanto menores são os trechos do sinal de voz permutados (ou volume deles) e maior é o número de opções possíveis de permutação deles, maior será a segurança das conversões, isto é tanto mais difícil recuperar o sinal de voz original e compreender o significado do discurso.

2. Segurança das conversões de encriptação

Esta seção contém informações gerais sobre a segurança das conversões.

2.1. Conceitos gerais da segurança das conversões de encriptação. Características quantitativas da segurança.

Conversões mosaicas de encriptação reorganizam os pequenos trechos do sinal de voz nos domínios da freqüência e do tempo.

A segurança criptográfica das conversões mosaicas de encriptação é caracterizada por vários parâmetros quantitativos.

2.1.1 Camuflagem de voz

Algumas conversões "reorganizam" os trechos do sinal de voz de forma ineficaz e para compreender o significado do que foi dito é suficiente simplesmente ouvir com atenção o discurso encriptado uma ou várias vezes. Neste caso, falamos de inteligibilidade residual alta ou camuflagem de voz baixa. Conversões seguras de encriptação são caracterizadas pela baixa inteligibilidade residual praticamente zerada.

2.1.2. Inteligibilidade da voz decriptada com uma chave errada

Se o intruso tiver um codificador de voz semelhante, ele pode interceptar o sinal e ouvir a voz desencriptada. No entanto, para fazer isso ele precisa saber a chave. Se digitar uma chave errada a desencriptação será realizada de maneira errada e em vez do discurso original será reproduzido um discurso com trechos rearranjados. Na

realidade se as conversões forem pouco seguras os intrusos ainda conseguem compreender palavras individuais. Conversões seguras são caracterizadas pela inteligibilidade baixa quando escutadas usando uma chave falsa ocasional já que os trechos do sinal de voz são misturados duas vezes.

2.1.3. Dimensões dos trechos do sinal de voz reorganizados

As dimensões são caracterizadas pelo comprimento dos trechos e pela largura da faixa de freqüências que eles usam. Para aumentar a segurança criptográfica estas dimensões devem ser as menores possíveis, mas isto costuma piorar a qualidade da voz recuperada. Por isso estas dimensões são seleccionadas de maneira que permita manter o balanço entre a segurança necessária e a qualidade da voz. Geralmente o comprimento dos trechos permutados não excede o comprimento do som da voz (100 msec) e a largura da faixa de freqüências destes trechos pode ser de 300 a 2000 Hz, isto é uma parte de um espectro do sinal de voz que não é inferior a 3 kHz. O volume do trecho do sinal de voz pode servir de uma característica generalizada do sinal de voz, isto é o resultado de multiplicação do comprimento pela largura da faixa de freqüências ocupada. Se os trechos permutados divergem em relação a esses parâmetros podemos considerar o volume médio.

2.1.4. Complexidade das conversões de encriptação

A complexidade das conversões de encriptação afeta diretamente a sua segurança e é determinada por algumas características qualitativas e quantitativas.

O número de opções para encriptar 1 segundo do sinal de voz é a característica quantitativa básica da complexidade de conversões. Para descobrir o número de possíveis opções de encriptação precisamos calcular a taxa de consumo da gama de encriptação por unidade de tempo.

Entretanto, este número não influencia diretamente a segurança das conversões.

Transformação de tempo-freqüência (TTF) – a conversão de um sinal de voz em que o sinal é comprimido no domínio do tempo e é esticado no domínio da

freqüência ou vice-versa. Em modo de recepção do sinal a operação é inversa. Com uma forte transformação do sinal de voz (de 2 a 4 vezes) a sua legibilidade é significativamente reduzida e a capacidade de reconhecer o locutor é eliminada.

Estrutura de quadro e escala móvel. Para rearranjar os trechos do sinal no domínio do tempo existem dois métodos diferentes. No método chamado de "permutação da estrutura de quadro" todo o sinal é dividido em quadros de duração fixa e a permutação é realizada dentro de um quadro. Assim é possível tentar todas as permutações possíveis dentro do quadro, o que é um ponto fraco deste tipo de permutações. O segundo método é chamado de "escala móvel" onde não é possível separar tais quadros e isto complica significativamente a recuperação do sinal de voz encriptado devido ao aumento da quantidade de tentativas das permutações possíveis.

Durações do trecho do sinal permutado. Os trechos do sinal de voz rearranjados podem ter as mesmas ou diferentes durações. No primeiro caso é bastante fácil realizar uma setorização geral do sinal encriptado marcando os limites dos trechos permutados, já que tal setorização neste caso será periódica. No segundo caso para realizar desencriptação é necessário definir os limites de cada um dos trechos permutados individualmente que complica significativamente o processo de desencriptação e reduz a sua eficácia.

Camuflagem dos momentos de comutação para diferentes durações de trechos cria as dificuldades adicionais durante desencriptação uma vez que isto reduz a confiabilidade da identificação das verdadeiras opções de permutação. Tal camuflagem pode ser realizada pelo equilibrio dos pontos de transição entre os trechos adjacentes do sinal de voz encriptado. No entanto, previamente tal operação não era usada em codificadores conhecidos e não foram realizados os estudos do seu impacto à segurança.

2.1.5. Custo de mão-de-obra aplicada para a desencriptação (restauração do sinal de voz encriptado)

Desde que em trechos rearranjados a estrutura do sinal de voz permanece sem alterações há uma possibilidade teórica de desencriptação (recuperação do discurso encriptado), no entanto quanto mais complexas são as conversões, mais habilidades

você precisa ter e mais trabalho deve ser aplicado. Para restaurar as conversões mais complexas é preciso ter uma qualificação profissional alta, nível de serviços considerado especializado nos países desenvolvidos, ter um equipamento adequado e gastar algumas horas para restaurar um segundo do discurso.

Avaliação do custo de mão-de-obra necessário para desencriptar a unidade de tempo do sinal de voz é uma característica básica da segurança.

No entanto, a obtenção de estimativas de custos de mão-de-obra para a desencriptação é um problema complexo, a própria solução empregada na sua totalidade é bastante trabalhosa, uma vez que a mesma implica na elaboração e a implementação do método de desencriptação e a recuperação do descurso de teste com a estimativa do tempo necessário para recuperar o significado da mensagem. Por isso o método analógico descrito na próxima seção é mais utilizado.

2.2. Método analógico na avaliação da segurança

O método mais simples que permite avaliar a segurança das conversões de voz é a comparação com as conversões conhecida a segurança das quais já foi avaliada. Para fazer isso é preciso selecionar o análogo mais próximo das conversões a segurança das quais já foi avaliada, identificar as diferenças entre as conversões analizadas e seu análogo e avaliar o impacto sobre a segurança de conversões.

3. Avaliação da segurança das conversões de encriptação desenvolvidas

3.1. Comprovação da segurança criptográfica das conversões de voz de encriptação SCR4

3.1.1. Características Gerais

As conversões de encriptação SCR4 são permutações mosaicas de 4 faixas de tempo-freqüência com uma comutação múltipla, uma transformação de tempo-freqüência que tem o factor constante e a chamada escala móvel.

3.1.2. Descrição das conversões

Depois de uma pré-filtragem com a freqüência baixa de corte de 3 kHz o sinal de voz que entra no codificador é digitalizado com a freqüência de amostragem de 6 kHz. Por meio do filtro de freqüência baixa com a freqüência de corte de 1500 Hz o sinal de entrada é dividido pelo filtro em duas faixas, desde que o sinal da parte inferior do espectro fica na posição direita e o sinal da parte superior do espectro é invertido em relação à freqüência de 1500 Hz. Assim as duas faixas de freqüências do sinal aparecem na faixa de freqüências de 0 a 1500 Hz. A dizimação do sinal é realizada duas vezes (afinamento de contagens), isto é de cada duas contagens consecutivas apenas 1 continua, o que leva à compressão do sinal no domínio do tempo e ao alongamento no domínio da freqüência por duas vezes. O sinal de cada um dos semi-canais sofre repetidamente a mesma operação, consequentemente o sinal inicial é dividido pelo filtro em 4 faixas de freqüências com a largura de 750 Hz e o sinal em cada faixa é comprimido no domínio do tempo e alongado no domínio da freqüência por 4 vezes, isto é, ocupa a faixa de freqüências de 3 kHz. Os sinais de cada uma das quatro faixas de freqüências são divididos em trechos de duração de 60 ms, que estão registrados nos buffers (retentores) de atraso. Para cada um dos subcanais estão destinados 2 buffers. O procedimento da leitura desses buffers é definido pela gama de encriptação. Ao mesmo tempo é feita permutação no domínio do tempo nos trechos dos sinais de subcanais. O trecho do sinal de sub-canal lido a partir do buffer sai do codificador de voz e é transmitido ao canal de comunicação. No lado do receptor a permutação inversa do trecho do sinal de voz é feita com a transformação simultânea de tempo-freqüência.

3.1.3. Características quantitativas da segurança SCR4

3.1.3.1. Camuflagem de voz

Devido a transformação de quatro vezes de tempo-freqüência o sinal encriptado parece com o canto de um pássaro por causa do aumento da freqüência do tom principal do sinal de voz em quatro vezes. Neste caso é totalmente impossível

compreender o sentido e reconhecer o locutor ou seja, a voz encriptada não contém os atributos de personalidade e de gênero.

Para ilustrar a conversão apresentamos o espectrograma do sinal original de voz (Fig. 1) e do sinal encriptado (Fig. 2).

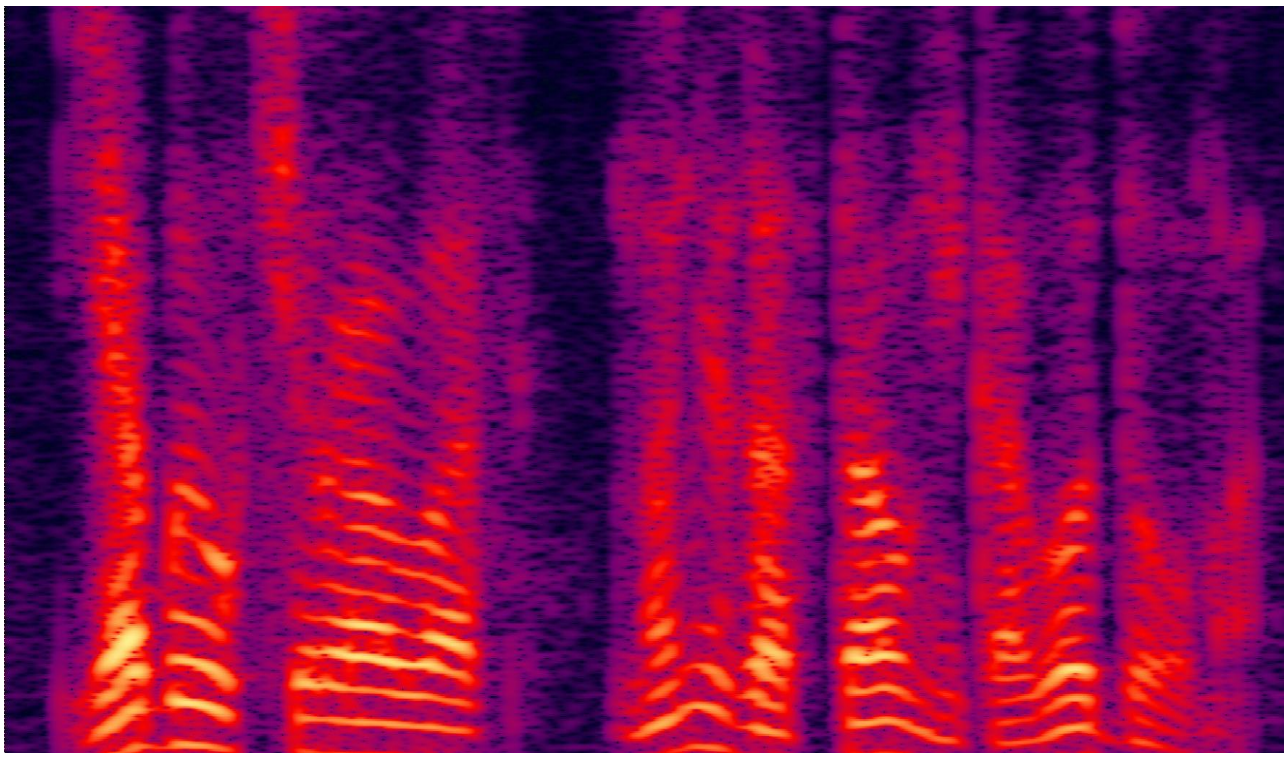

Fig. 1. Espectrograma do sinal original de voz.

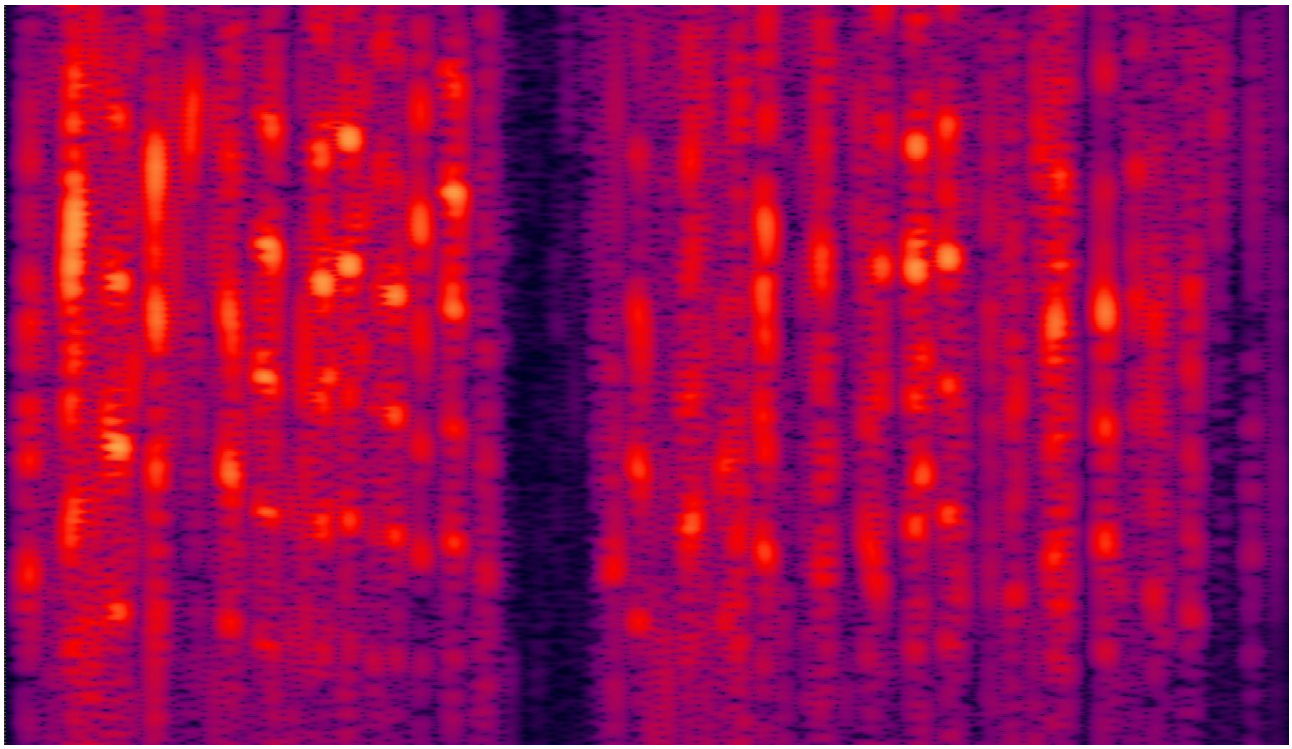

Fig. 2. Espectrograma do sinal encriptado por meio do codificador SCR4.

3.1.3.2. A quantidade de opções de encriptação

Um trecho mínimo do sinal de voz com uma duração de 60 milissegundos e uma largura da faixa de 750 Hz pode obter 5 opções de atraso no domínio do tempo e ocupar uma das quatro posições possíveis no domínio da freqüência. Neste caso 12 símbolos da gama binária são usados durante a encriptação destes 4 trechos. Assim, a encriptação de 1 segundo do sinal de voz consome 12/0,06 = 200 símbolos binários da gama. Com base nisso, o número total de opções para encriptar 1 segundo de discurso é $2^{200}=10^{60}$.

3.1.3.3. O volume do trecho do sinal de voz.

O comprimento do trecho é Δt=60 milissegundos, e a largura da faixa de freqüências Δf= 750 Hz. Assim, o volume do trecho do sinal de voz para as conversões de encriptação é:

$$V=\Delta t*\Delta f=0,06*750=45$$

3.3.1.4. Escuta com uma chave errada.

Segundo as estimativas dos especialistas, durante a escuta a inteligibilidade é muito baixa . O espectograma do sinal encriptado por meio do codificador SCR4 e desencriptado com uma chave errada é apresentada na Figura 3. Pode ser observado que os trechos do sinal de voz estão misturados nos domínios da freqüência e do tempo.

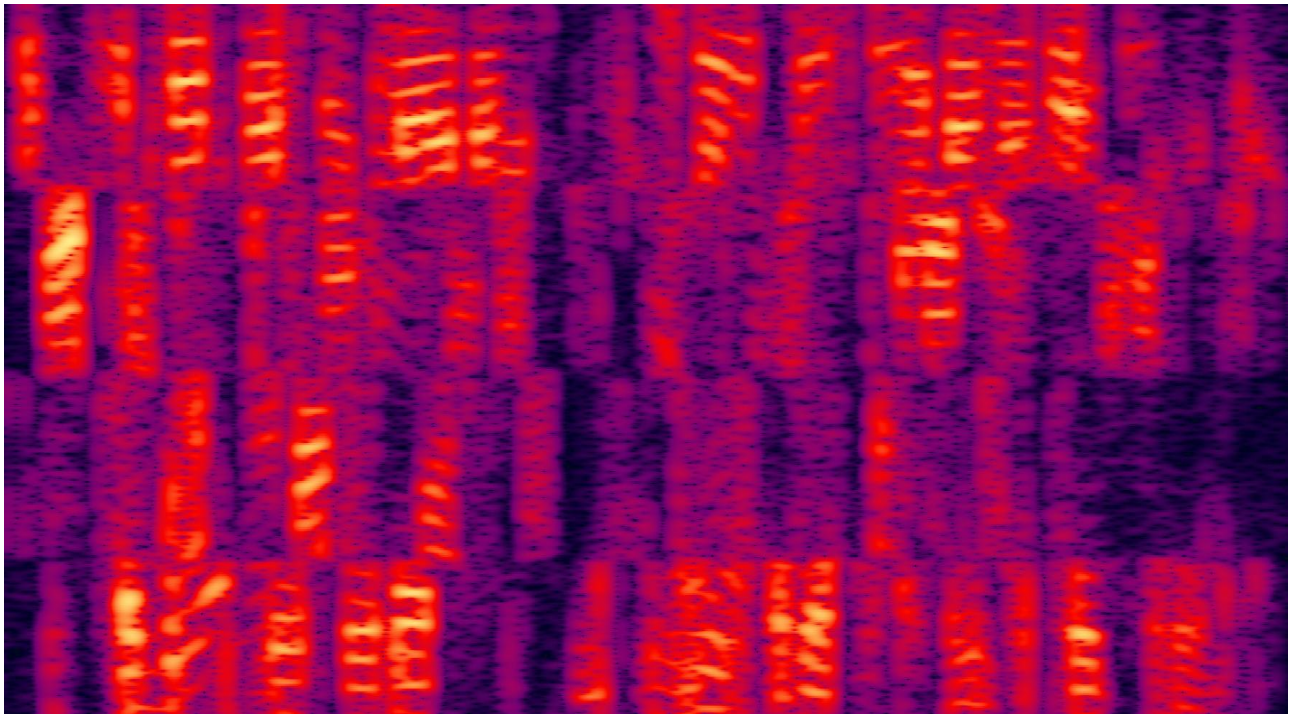

Fig. 3. O espectrograma do sinal encriptado por meio do codificador SCR4 e desencriptado com uma chave errada.

3.1.4. Análogos

Os princípios e parâmetros das conversões SCR4 (o comprimento e a faixa de freqüências dos trechos de voz não comutados, a taxa de consumo da gama de encriptação, a quantidade das opções possíveis de encriptação de 1 segundo de voz, escala móvel, comutação múltipla) são semelhantes às conversões de encriptação conhecidas a alta segurança das quais havia sido comprovada anteriormente.

A diferença está na presença da transformação de tempo-freqüência no SCR4 que aumenta o nível da camuflagem de voz no sinal encriptado em comparação com as conversões conhecidas.

3.2. Comprovação da segurança criptográfica das conversões de encriptação de voz baseadas nas permutações da faixa única no domínio do tempo com comutação aliquanta

3.2.1. Característica geral das conversões

As conversões realizam permutações no domínio do tempo com uma "escala móvel", durações variáveis dos trechos rearranjados do sinal e camuflagem dos limites dos trechos rearranjados.

3.2.2. Descrição das conversões

O sinal de voz recebido na entrada é dividido em trechos de durações variáveis que são definidas pela gama de encriptação. As durações dos trechos variam entre 20 e 70 ms, a duração média é de cerca de 45 ms. 3 ou 4 símbolos binários de gama são usados para desenvolver o valor da duração.

A permutação de trechos é realizada utilizando um buffer de atraso na faixa total de freqüências ocupada pelo sinal, que é um parâmetro de conversão e a largura da faixa de freqüências é de 2,7 a 3,5 kHz.

O procedimento de selecção de trechos do buffer é definido pela gama. 3 ou 4 símbolos da gama são usados para selecionar cada trecho. Neste caso o trecho pode ser invertido no domínio do tempo ou pode ser lido na posição inicial, dependendo do símbolo correspondente da gama. Assim, entre 7 e 9 símbolos binários da gama são usados para encriptar um trecho do sinal de voz com um comprimento médio de 45 milissegundos.

3.2.3. Características da complexidade das conversões

3.2.3.1. A quantidade de opções possíveis de encriptação de 1 segundo de voz.

Com base na descrição das conversões de 7 a 9 símbolos são usados para encriptar um trecho do sinal de voz. Considerando que o comprimento médio do trecho é 45 ms para encriptar 1 segundo de voz são usados

N=(7 …9)/0,045=156…200 bits da gama do codificador.

Assim, a quantidade das opções possíveis para encriptar 1 segundo do sinal de voz é $2^{156} = 6{,}4*10^{46}$ a $2^{200}=10^{60}$.

3.2.3.2. As características qualitativas da complexidade das conversões.

Diferentes comprimentos dos trechos do sinal de voz permutados juntamente com a escala móvel e a camuflagem dos limites dos trechos permutados caracterizam essa conversão como uma conversão altamente complicada. No entanto, o fato que as permutações são realizadas na faixa completa de freqüências do sinal de voz representa uma certa fraqueza que se manifesta em especial na camuflagem mais baixa da voz no sinal encriptado.

3.2.3.3. Camuflagem de voz.

De acordo com as estimativas preliminares, o locutor pode ser reconhecido e as vezes parece que é possível ouvir palavras familiares. No entanto, esta impressão pode ocorrer devido a semelhança do sinal encriptado com a fala. Por isso para uma avaliação fiável da camuflagem de voz é necessário realizar medições de articulação compostas da avaliação de peritos da inteligibilidade residual de voz de acordo com os resultados da escuta do sinal de voz encriptado pelos especialistas. O espectograma do sinal encriptado é apresentada na Fig. 4. Como pode ser visto, o espectrograma está semelhante ao espectrograma do sinal de voz (veja Fig. 1), o que também comprova que o sinal de voz encriptado por meio destas conversões é parecido com a fala.

Fig. 4. Espectrograma do sinal encriptado por meio das conversões da faixa única.

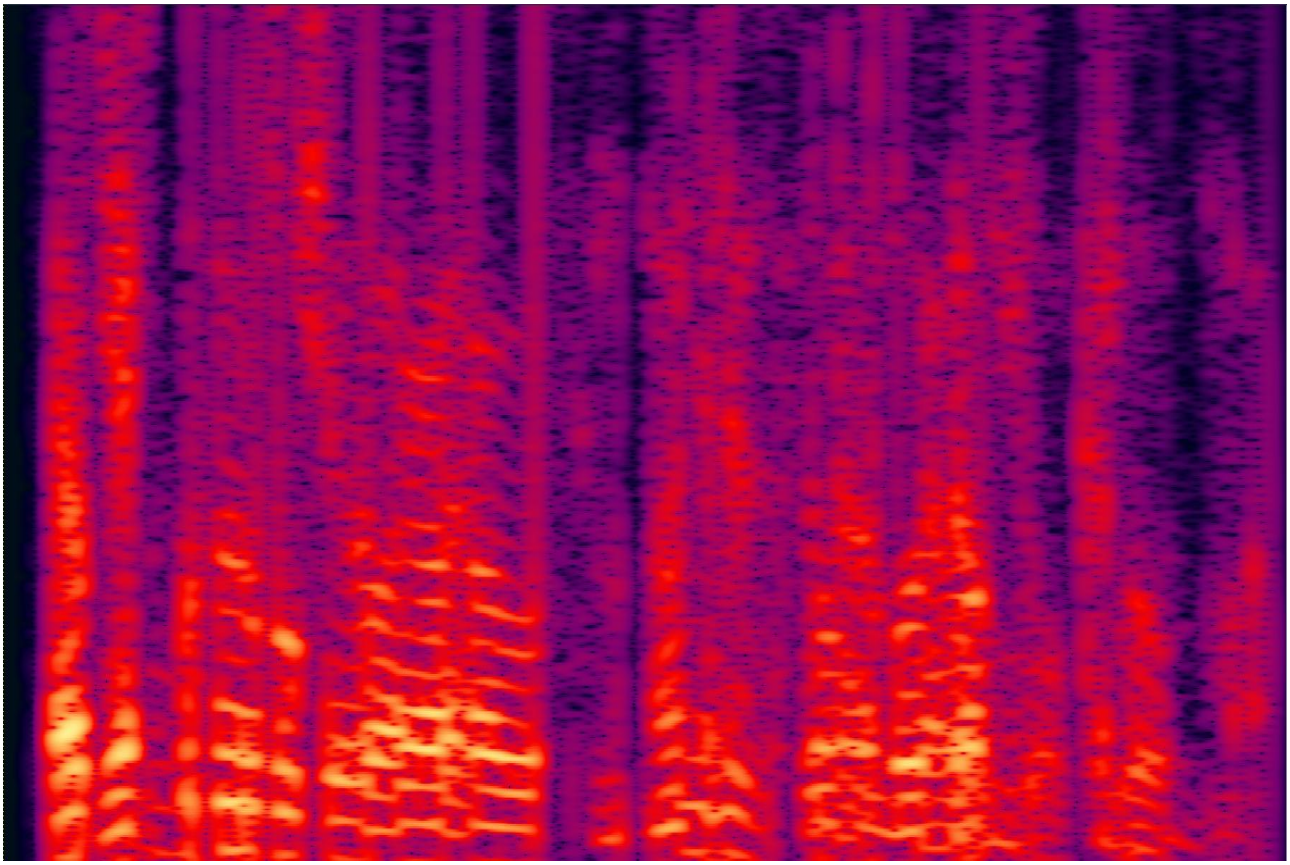

Fig. 4. Espectrograma do sinal encriptado por meio das conversões da faixa única.

3.1.3.4. O volume médio dos trechos do sinal de voz permutados.

O comprimento médio do trecho é Δt=45 ms, e a largura da faixa de freqüências é aproximadamente igual a Δf= 3000 Hz. Assim, o volume médio dos trechos do sinal de voz permutados para as conversões de encriptação analisadas é igual a

$$V=\Delta t*\Delta f=0{,}045*3000=135,$$

que é várias vezes maior do que o volume para as conversões SCR4, o que indiretamente caracteriza essas conversões como menos confiáveis.

3.2.4. Análogos

O análogo mais próximo é um equipamento conhecido, o esquema criptografado dele realiza as permutações de dois canais de tempo-freqüência com uma escala móvel e durações variádas de trechos permutados.

Sabe-se que tais conversões têm a segurança relativamente alta e não podem

ser desencriptadas automaticamente.

4. Sumário

O codificador executa a conversão dinâmica de tempo-freqüência de cada um dos trechos do sinal de voz com a permutação subsequente dos sinais no domínio do tempo.

A escuta do sinal encriptado não permite restaurar o significado das palavras ou conhecer o locutor.

A encriptação é realizada através de uma reorganização dos trechos de voz em tempo com conversões espectrais adicionais. Uma característica do algoritmo de encriptação é a incapacidade de desencriptar automaticamente o sinal de voz encriptado por meio de tentativas para descobrir a chave.