

# Nonlinear Spectral Subtraction Based on Tsallis Statistics for Speech Enhancement

## Nonlinear Spectral Subtraction Berbasis Tsallis Statistics untuk Peningkatan Kualitas Sinyal Ucapan

Hilman F. Pardede\*

Pusat Penelitian Informatika  
Lembaga Ilmu Pengetahuan Indonesia  
Komplek LIPI, Jalan Cisit No 21/154D Bandung 40135  
Jawa Barat, Indonesia

---

### Abstract

The presence of the noise degrades the quality and the intelligibility of the speech signal and hence reduces the performance of speech based application. Spectral subtraction is a popular method to remove additive noise. However, it has a major shortcoming of introducing musical noise. Several variants of spectral subtraction have been proposed to tackle this issue. One of them is the introduction of oversubtraction factor in the spectral subtraction formula. This approach nonlinear spectral subtraction. However, this factor is decided heuristically. Tsallis statistics has found to introduce a nonlinear subtraction naturally. A new variant of spectral subtraction, which is called  $q$ -spectral subtraction, has been derived.  $q$ -SS has been found to be effective for improving the robustness of speech recognition performance against noise. However, the evaluation of this method for speech enhancement tasks has not been explored yet. In this paper, the performance of  $q$ -spectral subtraction for speech enhancement task is investigated. It is found that  $q$ -SS is better than other spectral subtraction methods in improving the quality of speech signals.

**Key Words:** *speech enhancement, spectral subtraction, nonlinear spectral subtraction, musical noise*

### Abstrak

Adanya derau (*noise*) mengurangi kualitas dan inteligibilitas dari sinyal ucapan dan ini berakibat menurunnya performa dari aplikasi berbasis sinyal ucapan. Pengurangan spektral (*spectral subtraction*) adalah salah satu metode yang populer untuk menghilangkan derau tersebut. Akan tetapi, pengurangan spektral memiliki kelemahan, yaitu memperkenalkan musical noise. Telah banyak turunan dari pengurangan spektral yang diusulkan untuk mengurangi musical noise. Salah satunya adalah menggunakan oversubtraction dalam formulasi pengurangan spektral. Pendekatan ini disebut nonlinear pengurangan spektral. Akan tetapi, penentuan faktor ini secara heuristik. Dengan menggunakan Tsallis statistics, nonlinear subtraksi dapat diturunkan secara matematis. Varian baru *spectral subtraction* yang disebut  $q$ -spectral subtraction telah diturunkan. Metode ini telah terbukti efektif untuk meningkatkan performa sistem pengenalan ucapan terhadap noise. Akan tetapi, evaluasi metode ini untuk meningkatkan kualitas sinyal ucapan pada *speech enhancement* belum diinvestigasi. Pada paper ini, performa  $q$ -SS untuk *speech enhancement* akan diinvestigasi. Dari hasil percobaan, ditemukan bahwa  $q$ -SS lebih baik dalam meningkatkan kualitas sinyal ucapan dibandingkan metode pengurangan spektral lain.

**Kata kunci:** *speech enhancement, spectral subtraction, nonlinear spectral subtraction, derau musikal*

---

## 1. PENDAHULUAN

Derau (*noise*) dapat mengganggu dan merusak aspek persepsi suatu sinyal ucapan. Dengan kata lain, derau menurunkan kualitas (nyaman atau tidaknya suatu ucapan didengar) dan inteligibilitas (mudah atau tidaknya suatu ucapan dimengerti). *Speech enhancement* adalah suatu area penelitian yang bertujuan untuk meningkatkan kualitas dan inteligibilitas dari sinyal ucapan yang

---

\*Corresponding Author. Tel:

Email: hilm001@lipi.go.id

Received: 1 May 2013; revised: 14 May 2013; accepted: 14 May 2013

Published online: 22 May 2013

Corresponding editor: Bambang Sugiarto

(bambang@informatika.lipi.go.id)

© 2013 INKOM 2013/13-NO228

terkontaminasi oleh derau. Oleh karena derau dapat timbul di hampir semua kondisi, aplikasi *speech enhancement* menjadi sangat penting dan digunakan di banyak area terutama ketika suatu sistem diharapkan memiliki kinerja yang tahan (*robust*) terhadap derau. Contoh aplikasi *speech enhancement* misalnya digunakan pada tahap pra-prosesi pada sistem pengenalan ucapan (*speech recognition*), pada sistem komunikasi misalnya untuk aplikasi militer dimana komunikasi sering dilakukan pada keadaan yang banyak terdapat derau akibat baling-baling helikopter, suara senjata, dan lain-lain, untuk komunikasi *handsfree* yang kini semakin umum, dan untuk bidang kesehatan, *speech enhancement* digunakan pada alat bantu pendengaran untuk menghilangkan derau yang tidak diinginkan sebelum melakukan penguatan terhadap sinyal ucapan.

Mengingat pentingnya *speech enhancement*, banyak penelitian telah dilakukan dalam beberapa dekade terakhir. Lim dan Oppenheim [1] mengelasifikasikan *speech enhancement* kedalam tiga kategori. Kategori pertama adalah *speech enhancement* berbasis estimasi amplitudo dari spektral. Metode-metode yang termasuk pada kategori ini menggunakan short time spectral amplitude (STSA) untuk menghilangkan sinyal-sinyal yang tidak diinginkan untuk memperoleh estimasi dari sinyal “bersih” (*clean speech*) atau sinyal sebelum terkontaminasi derau. *Spectral subtraction* [2] adalah salah satu contoh dalam kategori ini. Kategori kedua adalah *speech enhancement* berbasis periodisitas dari sinyal ucapan. Metode-metode yang termasuk pada kategori ini bekerja berdasarkan karakteristik sinyal ucapan *voiced* (kondisi terdapat ucapan pada sinyal) yang dapat dibangun (disintesis) dari sinyal-sinyal periodik dengan frekuensi yang sama dengan frekuensi dasar (*fundamental frequency*) sinyal tersebut. Performa metode-metode pada kategori ini bergantung kepada tingkat akurasi dari estimasi periode *pitch* sinyal karena frekuensi dasar suatu sinyal ditentukan oleh periode dari *pitch* suatu sinyal. *Adaptive comb filter* [3], *harmonic selection* [1], dan *adaptive noise cancellation* [4] adalah beberapa contoh metode-metode pada kategori ini. Kategori terakhir adalah *speech enhancement* berbasis kepada model dari sinyal ucapan. *Speech enhancement* berbasis model statistika termasuk dalam kategori ini [5, 6]. Contoh lain adalah *speech enhancement* menggunakan *autoregressive* (AR) [7], *moving average* (MA) [8] dan *autoregressive moving average* (ARMA) [9]. Pada kategori ini sinyal ucapan dimodelkan. Kemudian, estimasi sinyal ucapan bersih dihasilkan menggunakan model tersebut.

*Spectral subtraction* merupakan salah satu metode yang paling populer untuk *speech enhancement*. Metode ini sederhana, mudah diimplementasikan, serta terbukti efektif mengurangi tingkat kekuatan derau. Akan tetapi, metode ini memiliki kelemahan-kelemahan, antara lain: metode ini memperkenalkan apa yang disebut derau musikal (*musical noise*). Derau musikal ini bahkan dapat lebih mengganggu dibandingkan derau asal. Derau musikal terjadi akibat tidak akuratnya estimasi dari spektrum derau. Hal ini berakibat terjadinya puncak yang terisolasi (*isolated peak*) pada spektrum. Puncak yang terisolasi ini ketika ditransformasikan ke domain waktu menjadi nada yang sangat mengganggu pendengar. Usaha mengurangi derau musikal merupakan salah satu fokus penelitian pada *spectral subtraction*. Telah banyak usaha penelitian dilakukan untuk mengurangi efek derau musikal ini. Secara teori, derau ini dapat dihilangkan bila estimasi derau lebih akurat. Akan tetapi, derau pada umumnya adalah sinyal *nonstationary*, estimasi derau yang baik sangat sulit untuk diperoleh.

Berbagai pendekatan lain juga telah diusulkan untuk mengurangi efek derau musikal selain memperbaiki estimasi derau. Boll [2] mengusulkan agar setiap nilai negatif dari spektral akibat *spectral subtraction* digantikan oleh nilai minimum spektra dari beberapa frame sesudahnya. Performa yang lebih lebih baik diperoleh daripada membuat nilai tersebut menjadi nol. Akan tetapi, hal ini akan membuat *spectral subtraction* tidak *real time*. Berouti [10] memperkenalkan faktor *oversubtraction* dan *flooring* dalam *spectral subtraction*. Metode ini sering disebut *nonlinear spectral subtraction* (NSS). *Oversubtraction* (umumnya  $> 1$ ) digunakan sebagai faktor pengali pada estimasi derau. Tujuannya adalah untuk mengurangi besarnya amplitudo puncak spektrum dari *spectral subtraction*. Dengan demikian, tingkat kekerasan (*loudness*) dari derau musikal dapat ditekan. Dengan memberikan *flooring* untuk nilai negatif pada spektrum, beda ketinggian antara puncak spektrum yang terisolasi tersebut dengan spektrum sekitarnya juga dapat dikurangi. Banyak variasi spektral subtraction yang diturunkan dari formulasi Berouti [11–14].

Akan tetapi, pemberian faktor *oversubtraction* lebih merupakan pendekatan heuristik (*common sense*). Sehingga, belum ada metode untuk optimasi nilai faktor ini yang konsisten dan pada umumnya ditentukan secara empiris. Hilman et. al [15] menurunkan varian baru dalam *spectral subtraction* yang diturunkan dari *Tsallis statistics*. Metode ini, yang dinamakan *q-spectral subtraction* (*q-SS*), diturunkan dengan menggunakan asumsi bahwa sinyal ucapan yang terkontaminasi derau

(noisy) distribusi  $q$ -Gaussian. Dengan menggunakan model ini, formula baru yang juga dapat dilihat sebagai varian dari NSS dapat diturunkan secara matematis. Dengan penurunan ini, cara yang lebih konsisten untuk menentukan faktor pengali pada formula *spectral subtraction* yang optimum dapat diturunkan. Hasil penelitian pada sistem pengenalan ucapan (*speech recognition*) menunjukkan bahwa performa  $q$ -SS lebih baik daripada metode *spectral subtraction* yang konvensional. Namun, kinerja  $q$ -SS untuk peningkatan kualitas dan inteligibilitas sinyal ucapan belum dieksplorasi secara menyeluruh. Hal ini menjadi penting karena beberapa penelitian menunjukkan bahwa peningkatan performa sistem pengenalan ucapan tidak selalu berarti peningkatan kualitas sinyal ucapan itu sendiri.

Pada paper ini, kinerja  $q$ -SS dalam meningkatkan kualitas dan inteligibilitas sinyal ucapan yang terkontaminasi derau akan diinvestigasi. Hasil percobaan menunjukkan bahwa performa  $q$ -SS lebih baik daripada NSS secara umum untuk *speech enhancement*. Hasil percobaan juga mengindikasikan bahwa derau musikal lebih sedikit ketika menggunakan  $q$ -SS daripada NSS.

Struktur paper ini adalah sebagai berikut. Pada Bagian 2 akan dijelaskan mengenai metode *spectral subtraction* dan beberapa variasinya. Pada Bagian 3 akan dijelaskan secara lebih rinci mengenai *Tsallis statistics* dan metode yang diusulkan:  $q$ -spectral subtraction. Mengenai implementasi  $q$ -SS akan dijelaskan pada Bagian 4. Pada Bagian 5 dijelaskan bagaimana prosedur percobaan dilakukan dan cara mengevaluasi performa dari  $q$ -SS. Hasil percobaan akan dijelaskan dan didiskusikan pada Bagian 6. Bagian 7 merupakan kesimpulan dari paper ini.

## 2. OVERVIEW SPECTRAL SUBTRACTION

### 2.1 Formulasi

Metode *spectral subtraction* barangkali salah satu metode yang paling populer untuk menghilangkan derau pada sinyal ucapan. Pada metode ini, sinyal *enhanced* (sinyal hasil *spectral subtraction* atau sinyal yang telah dihilangkan tingkat derau) dilakukan dengan mengurangi spektrum sinyal *noisy* (sinyal yang terkontaminasi oleh derau) dengan estimasi sinyal derau. Jika sinyal bersih dinotasikan dengan  $x(t)$ , sinyal derau dinotasikan dengan  $n(t)$ , dan sinyal *noisy* dengan  $y(t)$ , dimana  $t$  merupakan waktu, maka secara matematis hubungan  $x(t)$ ,  $n(t)$ , dan  $y(t)$  pada domain waktu dapat dituliskan sebagai berikut:

$$y(t) = x(t) + n(t). \quad (1)$$

Dengan mengambil Discrete Fourier Transform (DFT), Persamaan (1) dapat direpresentasikan dalam

domain frekuensi menjadi:

$$Y(m, k) = X(m, k) + N(m, k), \quad (2)$$

dimana  $Y(m, k)$ ,  $X(m, k)$ , dan  $N(m, k)$  adalah representasi  $y(t)$ ,  $x(t)$ , dan  $n(t)$  pada domain frekuensi saat frame  $m$  dan indeks frekuensi  $k$ . Dengan mengambil nilai magnitudo Persamaan (2) dan menguadratkannya, Persamaan (2) dapat dituliskan dalam domain power spectral sebagai berikut:

$$\begin{aligned} |Y(m, k)|^2 &= |X(m, k)|^2 + |N(m, k)|^2 \\ &\quad + 2|X(m, k)||N(m, k)| \cos \theta_{XN}, \end{aligned} \quad (3)$$

dimana  $\theta_{X,N}$  adalah beda fasa antara  $X$  dan  $N$ . Karena nilai ekspektasi  $\mathbb{E}\{\cos \theta_{XN}\} = 0$ , maka  $2|X(m, k)||N(m, k)| \cos \theta_{XN} = 0$  dan Persamaan (3) dapat disederhanakan menjadi:

$$|Y(m, k)|^2 = |X(m, k)|^2 + |N(m, k)|^2. \quad (4)$$

Jadi, dari Persamaan (4), dapat dilihat bahwa bila spektrum derau  $N(m, k)$  dapat diestimasi, maka sinyal bersih  $X(m, k)$  dapat diperoleh dengan mengurangi spektrum derau dari spektrum sinyal *noisy*  $Y(m, k)$ . Maka *spectral subtraction* dapat diformulasikan [2]:

$$|\hat{X}(m, k)|^2 = |Y(m, k)|^2 - |\hat{N}(m, k)|^2, \quad (5)$$

Dimana  $|\hat{X}(m, k)|^2$  adalah estimasi spektrum power dari sinyal  $X$  atau disebut juga sinyal *enhanced* dan  $|\hat{N}(m, k)|^2$  adalah estimasi spektrum power dari derau. Karena Persamaan (5) dilakukan pada domain power spektral, metode ini juga sering disebut *power spectral subtraction* (PSS). *Spectral subtraction* juga dapat dilakukan pada domain magnitudo spectral selain domain power spectral. Dari sudut pandang statistik, formulasi *spectral subtraction* dapat diturunkan dengan memaksimalkan varians dari distribusi sinyal *noisy* dengan asumsi spektrum sinyal *noisy*  $Y$  terdistribusi secara Gaussian [16].

Akibat karakteristik sinyal derau yang *nonstationary*, sangat sulit memperoleh estimasi derau yang akurat. Oleh karena itu, akan selalu terdapat perbedaan antara nilai spektrum derau yang sesungguhnya dan nilai estimasinya. Hal ini berakibat dua hal: Pertama, ada derau yang tertinggal pada sinyal *enhanced* yang disebut *residual noise*. Ini terjadi apabila nilai aktual derau lebih besar dari nilai estimasinya. Kedua, terjadinya distorsi pada sinyal suara (*speech distortion*). Hal ini terjadi apabila nilai aktual derau lebih kecil dari estimasinya sehingga ada bagian sinyal ucapan yang akan hilang/terdistorsi. Selain itu, dapat dilihat dari Persamaan (5), ketika  $\hat{N}(m, k) > Y(m, k)$

maka akan menghasilkan nilai power yang negatif. Hal ini secara teoretis tidak boleh terjadi, sehingga untuk mengatasinya dapat dengan menolkan nilai tersebut atau melakukan *flooring* dengan berbagai teknik [17]. Akibat dari kelemahan-kelemahan ini, pada spektrum sinyal *enhanced*, dapat terjadi energi spektrum yang cukup besar yang terisolasi. Akibat adanya isolasi puncak (*peak*) pada spektrum, ketika sinyal tersebut ditransformasi ke domain waktu, puncak-puncak spektrum yang terisolasi tersebut menjadi bunyi *tone* (nada) yang sangat mengganggu yang dikenal menjadi derau musikal (*musical noise*). Usaha untuk menghilangkan derau musikal ini menjadi salah satu topik penelitian yang hangat pada *speech enhancement*.

## 2.2 Varian Spectral Subtraction

Salah satu teknik yang banyak diadopsi untuk mengurangi derau musikal adalah teknik yang diusulkan Berouti [10]. Pada teknik ini, derau musikal dapat dikurangi dengan memberi faktor pengali pada spektrum derau. Faktor ini yang disebut *oversubtraction*. Dengan memberi nilai Dengan demikian ketinggian peak yang terisolasi tersebut dapat dikurangi dan dengan sendirinya tingkat kenyaringan derau musikal juga akan berkurang. Dengan teknik ini, metode *spectral subtraction* menjadi ( $m$  dan  $k$  dihilangkan untuk simplifikasi notasi)

$$|\hat{X}|^2 = \begin{cases} |Y|^2 - \alpha|\hat{N}|^2 & \text{if } |Y|^2\alpha|\hat{N}|^2 > \beta|Y|^2, \\ \beta|Y|^2 & \text{else} \end{cases} \quad (6)$$

dimana  $\alpha$  adalah *oversubtraction*,  $\beta$  adalah faktor *flooring*, dan SNR adalah *signal-to-noise ratio*, yaitu ratio antara sinyal dan derau. Nilai  $\alpha$  umumnya ditentukan secara adaptif tergantung kepada nilai SNR nya. Dengan adanya faktor  $\alpha$  ini, *spectral subtraction* menjadi nonlinear. Oleh karena itu juga sering disebut Nonlinear Spectral Subtraction (NSS). Banyak varian dari NSS di literatur, antara lain [11–13]. Umumnya perbedaannya adalah bagaimana menentukan nilai  $\alpha$ . Berouti [10] menentukan relasi  $\alpha$  dan SNR untuk setiap frame  $m$  dan frekuensi  $k$  sebagai berikut:

$$\alpha(m, k) = \begin{cases} 1 & \text{if } \text{SNR}(m, k) \geq 20\text{dB}, \\ 4 - \frac{3}{20}\text{SNR}(m, k) & \text{if } -5\text{dB} \leq \text{SNR}(m, k) < 20\text{dB}, \\ 4.75 & \text{if } \text{SNR}(m, k) < -5\text{dB}, \end{cases} \quad (7)$$

SNR dihitung untuk tiap frekuensi dan frame sebagai berikut:

$$\text{SNR}(m, k) = 10 \log \frac{|Y(m, k)|^2}{|\hat{N}(m, k)|^2}. \quad (8)$$

Akan tetapi pemberian faktor  $\alpha$  lebih merupakan pendekatan heuristik (*common sense*). Oleh karena itu tidak ada satu metode konsisten bagaimana penentuan nilai optimum dari  $\alpha$ . Pendekatan matematis dan lebih natural untuk varian *spectral subtraction* dapat diturunkan dengan menggunakan konsep *Tsallis statistics* pada bidang mekanika statistik. Dengan konsep ini, sebuah teknik baru, yang juga dapat dikategorikan sebagai varian NSS telah diturunkan. Metode ini dinamakan  $q$ -spectral subtraction ( $q$ -SS) [15] dan akan dijelaskan pada Bagian 3.

## 3. METODE

### 3.1 Tsallis Statistics

Konsep *non-extensive statistics* diperkenalkan oleh Tsallis [18]. Teori ini digunakan untuk menggeneralisasikan Shannon *entropy* dan Boltzmann-Gibbs *statistics*. Konsep ini menggunakan fungsi *q*ekponensial ( $q$ -exp) dan  $q$ -logaritma ( $q$ -log) yang didefinisikan sebagai berikut:

$$\exp_q(x) = (1 + (1 - q)x)^{\frac{1}{1-q}}, \quad (9)$$

$$\log_q(x) = \frac{x^{1-q} - 1}{1 - q}. \quad (10)$$

Fungsi-fungsi ini adalah salah satu generalisasi fungsi eksponensial dan logaritma berbasis  $e$  (*natural logarithm*) ketika  $q$  mendekati 1. Akan tetapi, berbeda dengan fungsi eksponensial dan logaritma yang merupakan fungsi *homomorphic* dan *extensive*, fungsi-fungsi ini adalah fungsi *non-extensive* ketika  $q \neq 1$  [19]. Dalam teori ini, Tsallis *entropy* didefinisikan sebagai berikut:

$$S_q = -k \int p_i(x) \log_q p_i(x). \quad (11)$$

Dengan menggunakan fungsi  $q$ -log, Tsallis *entropy* menjadi *non-extensive* dan akan sama dengan Shannon *entropy* ketika  $q = 1$ . Sehingga Tsallis *entropy* merupakan salah satu generalisasi Shannon *entropy*. Konsep Tsallis ini diturunkan untuk menjelaskan fenomena *non-extensive* pada berbagai sistem kompleks. Dengan konsep ini paramete  $q$  digunakan untuk menjelaskan tingkat kekompleksan suatu sistem. Lebih jelas mengenai konsep Tsallis *entropy* dan Tsallis *statistics* dapat dilihat di [20,21].

Dalam Tsallis *statistics*, distribusi  $q$ -Gaussian dapat diturunkan. Distribusi ini diperoleh dengan

memaksimumkan Tsallis *entropy* dengan cara yang sama distribusi Gaussian dapat diturunkan dari Shannon *entropy*. Distribusi  $q$ -Gaussian adalah distribusi *heavy-tailed* ketika  $q > 1$ . Distribusi yang *heavy-tailed* adalah distribusi dimana ekor (*tail*) distribusi lebih besar dibandingkan distribusi Gaussian. Banyak sistem kompleks didapati memiliki tipe distribusi seperti ini. Distribusi  $q$ -Gaussian dengan nilai rata-rata nol dan varians  $\lambda_q$  is definisikan sebagai berikut:

$$P_q(X) = \frac{A_q B_q}{\sqrt{\lambda_q}} \exp_q \left( -\frac{B_q^2 |X|^2}{\lambda_q} \right), \quad (12)$$

dimana  $A_q$  adalah faktor normalisasi yang diformulasikan sebagai berikut:

$$A_q = \begin{cases} \frac{\Gamma(\frac{5-3q}{2-2q})}{\Gamma(\frac{2-q}{1-q})} \sqrt{\frac{1-q}{\pi}} & -\infty < q < 1 \\ \frac{1}{\sqrt{\pi}} & q = 1 \\ \frac{\Gamma(\frac{1}{q-1})}{\Gamma(\frac{3-q}{2q-2})} \sqrt{\frac{q-1}{\pi}} & 1 < q < 3, \end{cases} \quad (13)$$

dan  $B_q$  adalah faktor skala  $B_q = \frac{1}{\sqrt{3-q}}$ . Untuk lebih detail mengenai distribusi  $q$ -Gaussian, dapat mengacu kepada [22].

### 3.2 Formulasi $q$ -Spectral subtraction

Sinyal ucapan dapat dikategorikan sebagai sistem kompleks [23]. Sinyal ucapan adalah sinyal *quasi-stationary* dan terdapat berbagai korelasi antar komponen penyusun sinyal suara secara kompleks dan memiliki skala waktu korelasi yang juga beragam. Oleh karena itu model Gaussian bukanlah model terbaik untuk sinyal suara seperti pada PSS. Dengan memodelkan sinyal noisy terdistribusi  $q$ -Gaussian,  $q$ -Spectral subtraction diformulasikan sebagai berikut:

$$|\hat{X}(m, k)|^2 = \frac{2(2-q)}{3-q} |Y_f|^2 - |N_f|^2. \quad (14)$$

Dari Persamaan (14) dapat dilihat,  $q$ -SS akan sama dengan Persamaan 5 ketika  $q = 1$ . Sehingga  $q$ -SS merupakan generalisasi dari PSS.

Pada [15], telah ditunjukkan bahwa parameter  $q$  pada  $q$ -SS dan parameter  $\alpha$  pada NSS memiliki relasi sebagai berikut:

$$\alpha = \frac{3-q}{2(2-q)}. \quad (15)$$

Sehingga dengan kata lain,  $q$ -SS dapat diadaptasi kepada berbagai varian NSS. Hubungan ini telah terbukti pada sistem pengenalan ucapan. Kedua metode ini,  $q$ -SS dan NSS, memiliki performa yang serupa. Hal ini dikarenakan pada sistem

pengenalan ucapan, kesamaan informasi statistik (nilai rata-rata dan varians distribusi) kondisi training dan testing lebih menentukan performa sistem pengenalan ucapan dan operasi skaling tidak akan mempengaruhi performa sistem pengolahan ucapan.

Berdasarkan Persamaan (15) dan (7), maka nilai optimum  $q$  dapat diformulasikan sebagai berikut:

$$\hat{q} = \begin{cases} 1 & \text{if } \text{SNR}(m, k) \geq 20\text{dB}, \\ 2 + \frac{10}{3\text{SNR}(m, k) - 70} & \text{if } -5\text{dB} \leq \text{SNR}(m, k) < 20\text{dB}, \\ 4.75 & \text{if } \text{SNR}(m, k) < -5\text{dB}, \end{cases} \quad (16)$$

### 3.3 Perbandingan $q$ -SS dan NSS

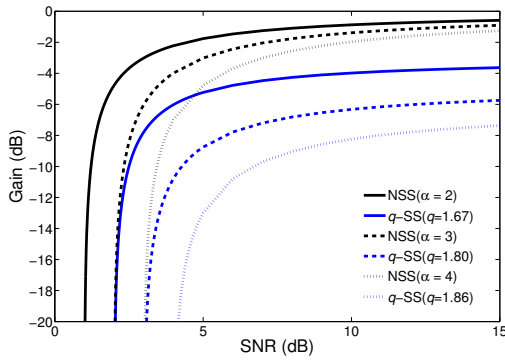
Meskipun  $q$ -SS dan NSS memiliki performa yang serupa ketika diimplementasikan pada sistem pengenalan ucapan, hal ini berbeda dengan implementasi pada *speech enhancement*. Pada *speech enhancement*, perbedaan spektrum sinyal *enhanced* memegang peranan dalam menentukan performa metode tersebut. Untuk mengamati perbedaan  $q$ -SS dan NSS untuk *speech enhancement* dapat dilihat dari kurva *Gain* kedua metode tersebut.

Kurva *Gain* digunakan untuk melihat sebesar apa suatu sinyal atau derau (dalam dB) ketika NSS dan  $q$ -SS diimplementasikan. Hal ini penting untuk mengamati seberapa besar derau dikurangi (*noise reduction*) dan seberapa besar sinyal ucapan terdistorsi (*speech distortion*). Pada *speech enhancement*, kedua faktor ini memegang peranan paling penting untuk menentukan kualitas dan inteligibilitas sinyal ucapan. Tujuan utama *speech enhancement* adalah *noise reduction* yang semaksimal mungkin dengan *speech distortion* yang seminimum mungkin. Namun, sering terjadi trade off antara *noise reduction* dan *speech distortion* pada *speech enhancement*. Ini akibat ketidak akuratan estimasi derau, pada umumnya ketika semakin besar *noise reduction* maka *speech distortion* juga semakin besar. Secara umum, ketika *Gain* semakin besar, maka artinya *noise reduction* semakin kecil dan *speech distortion* juga semakin kecil. Sebaliknya ketika *Gain* semakin kecil maka *noise reduction* akan tinggi tetapi *speech distortion* juga tinggi.

*Gain* dapat diperoleh dengan formula sebagai berikut:

$$\text{Gain} = 20 \log_{10} H, \quad (17)$$

dimana  $H$  adalah fungsi transfer suatu filter. Oleh karena itu, untuk memperoleh *Gain*,  $q$ -SS dan NSS harus direpresentasikan dalam bentuk fungsi transfer



Gambar 1: Komparasi kurva atenuasi  $q$ -SS dan NSS untuk beberapa nilai  $q$  dan nilai  $\alpha$  berdasarkan Persamaan 15.

(transfer function) sebagai berikut:

$$|\hat{X}(m, k)| = H_{q\text{-SS}}|Y(m, k)| \quad (18)$$

$$|\hat{X}(m, k)| = H_{\text{NSS}}|Y(m, k)| \quad (19)$$

dimana  $H_{q\text{-SS}}$  dan  $H_{\text{NSS}}$  adalah fungsi transfer  $q$ -SS dan NSS yang diformulasikan sebagai berikut:

$$H_{q\text{-SS}} = \left( \frac{v(q)|Y(m, k)|^2 - |N(m, k)|^2}{|Y(m, k)|^2} \right)^{0.5} \quad (20)$$

$$H_{\text{NSS}} = \left( \frac{|Y(m, k)|^2 - \alpha|N(m, k)|^2}{|Y(m, k)|^2} \right)^{0.5} \quad (21)$$

dimana  $v(q) = \frac{2(2-q)}{3-q}$ . Jika dinotasikan  $\gamma = \frac{|Y(m, k)|^2}{|N(m, k)|^2}$ , maka Persamaan. (20) dan (21) dapat dituliskan menjadi:

$$H_{q\text{-SS}} = \left( \frac{v(q) \cdot \gamma_f - 1}{\gamma} \right)^{0.5} \quad (22)$$

$$H_{\text{NSS}} = \left( \frac{\gamma - \alpha}{\gamma} \right)^{0.5} \quad (23)$$

Persamaan (23) dapat ditulis ulang menjadi:

$$H_{\text{NSS}} = \left( \frac{\alpha \left( \frac{\gamma}{\alpha} - 1 \right)}{\gamma} \right)^{(0.5)} \quad (24)$$

Berdasarkan Persamaan (22) dan (24), maka fungsi transfer  $q$ -SS dan NSS memiliki hubungan sebagai berikut:

$$H_{\text{NSS}} = \left( \frac{1}{v(q)} \right)^{0.5} H_{q\text{-SS}} \quad (25)$$

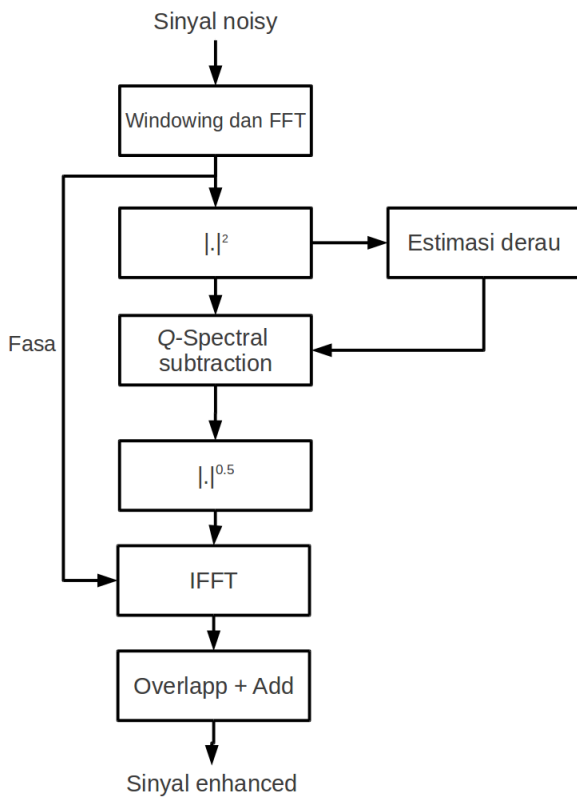
Ketika  $q > 1$  maka  $v(q) < 0$ . Maka dari Persamaan (25) kita dapat lihat bahwa  $Gain$  dari NSS lebih besar dari  $q$ -SS.

Gambar 1 membandingkan kurva  $Gain$  dari NSS dan  $q$ -SS sebagai fungsi dari SNR. Beberapa nilai  $\alpha$  dan  $q$  dibandingkan. Berdasarkan gambar tersebut dan Persamaan (25), dapat dilihat bahwa NSS memiliki  $Gain$  yang lebih tinggi dibandingkan  $q$ -SS ketika SNR tinggi. Hal ini berarti pada SNR yang tinggi, secara umum NSS akan mengurangi derau lebih kecil dibandingkan  $q$ -SS. Artinya  $q$ -SS lebih baik dalam mengurangi derau pada kondisi SNR yang tinggi. Selain itu, dari kurva tersebut dapat juga dilihat bahwa *flooring* terjadi pada SNR yang lebih tinggi pada  $q$ -SS dibandingkan NSS. Misalnya, untuk  $\alpha = 3$ , yaitu sama dengan  $q = 1.8$  (berdasarkan Persamaan (15)), *flooring* terjadi ketika SNR 2dB pada NSS dan 3dB pada  $q$ -SS. Hal ini menjadi penting karena secara umum mengestimasi derau ketika SNR rendah lebih sulit dibandingkan ketika SNR tinggi. Ini dikarenakan sulit melakukan deteksi ada tidaknya ucapan pada sinyal tersebut. Oleh karena itu kemungkinan terjadinya *speech distortion* menjadi lebih besar pada kondisi SNR rendah. Dengan melakukan *flooring* pada area ini dapat meminimalkan hal ini, dan dengan demikian dapat mengurangi terjadinya penurunan kualitas dan inteligibilitas sinyal ucapan. Dengan kata lain,  $q$ -SS memiliki potensi menghilangkan derau lebih baik pada SNR yang tinggi, dan dapat meminimumkan terjadinya distorsi pada sinyal ucapan ketika SNR rendah.

#### 4. IMPLEMENTASI

Implementasi  $q$ -SS untuk *speech enhancement* dapat dilihat pada Gambar 2. Proses *windowing* dilakukan pada sinyal ucapan menggunakan *Hamming window* dengan panjang *window* 20-ms dan *overlap* antara tiap frame sebesar 10-ms. Setiap frame dari sinyal ucapan kemudian ditransformasi ke domain frekuensi dengan Fast Fourier Transform (FFT) dengan total  $K = 256$  komponen frekuensi. Kemudian fasa dari FFT disimpan yang untuk kemudian digunakan sebagai fasa pada proses IFFT sinyal *enhanced*. Kemudian power sinyal tersebut diperoleh dengan mengambil nilai magnitude dari spektrum dan mengkuadratkannya. Setelah itu sinyal noisy dalam domain power spektral melalui proses  $q$ -SS untuk memperoleh sinyal *enhanced*. Kemudian fasa yang telah disimpan kemudian ditambahkan ke sinyal *enhanced* untuk kemudian dikonversi ke domain waktu dengan Inverse Fast Fourier Transform (IFFT).

Untuk estimasi derau pada paper ini, sebagai nilai awal, digunakan nilai rata2 sinyal noisy untuk 5 frame pertama. Hal ini dapat dilakukan dengan asumsi bahwa tidak ada ucapan pada awal dari sinyal, atau dengan kata lain hanya ada derau. Asumsi ini umum dilakukan untuk estimasi



Gambar 2: Diagram blok proses *speech enhancement* dengan *q-SS*.

noise pada *speech enhancement*. Untuk frame-frame berikutnya, mengingat noise bersifat nonstationary, maka estimasi derau perlu diperbaharui (*update*). Berikut adalah teknik yang digunakan untuk *update* estimasi derau:

- (1) Mengalkulasi SNR untuk tiap frame. SNR dapat dikalkulasi sebagai berikut:

$$SNR = 20 \log_{10} \frac{\sqrt{\sum_{k=0}^{K-1} |Y(m, k)|^2}}{\sqrt{\sum_{k=0}^{K-1} |2\hat{N}(m, k)|^2}}. \quad (26)$$

- (2) Membandingkan SNR dengan ambang batas yang ditentukan. Jika SNR pada frame tersebut dibawah ambang batas, maka diasumsikan bahwa tidak ada ucapan pada frame tersebut, dan dilakukan *update* estimasi derau. Hal ini dikarenakan ketika tidak ada ucapan, maka SNR pada frame tersebut akan rendah. *Update* dilakukan mengikuti formulasi sebagai berikut:

$$\hat{N}(m, k) = \begin{cases} \ddot{N}(m, k) & \text{if } SNR < T, \\ \hat{N}(m - 1, k) & \text{else,} \end{cases} \quad (27)$$

dimana  $T$  adalah nilai ambang batas SNR. Pada paper ini  $T = 3$ .  $\ddot{N}(m, k)$  diperoleh dengan

formula sebagai berikut:

$$\ddot{N}(m, k) = G|\hat{N}(m - 1, k)|^2 + (1 - G)|Y(m, k)|^2, \quad (28)$$

dimana  $G = 0.9$ .

Sementara itu, untuk *flooring*, digunakan  $\beta = 0.002$ .

## 5. PERENCANAAN PERCOBAAN

### 5.1 Speech Corpus

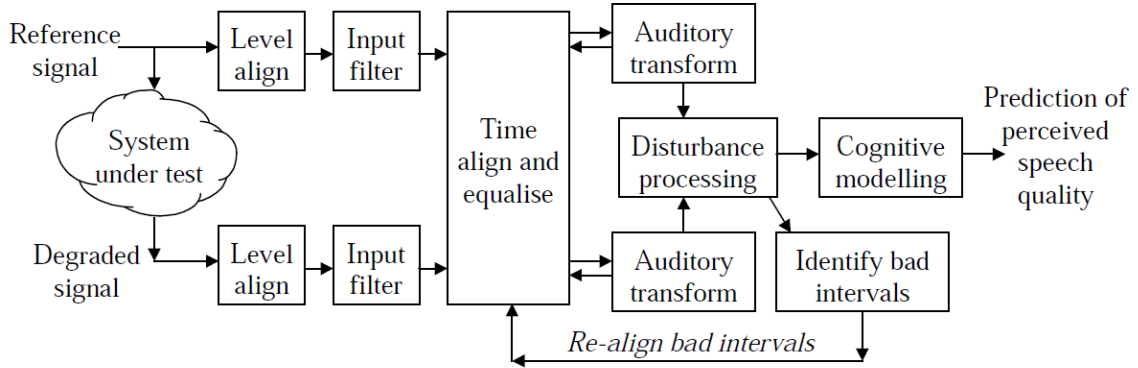
Pada penelitian ini, Noizeus *speech corpus* [24] digunakan untuk mengevaluasi performa *q-SS*. Noizeus terdiri dari 30 kalimat yang secara fonetis seimbang. Pembicara pada korpus ini terdiri dari tiga orang pria dan tiga orang wanita. Data suara ini memiliki frekuensi sampling 8 kHz. Setiap data dilakukan proses filtering untuk memperoleh karakteristik frekuensi dari *handset* telepon. Setiap sinyal pada Noizeus kemudian ditambahkan derau secara artifisial pada SNR 0 – 15 dB. Ada delapan jenis derau yang ditambahkan: *babble* (suara orang berbicara di keramaian), *airport* (suara keadaan di bandara), *exhibition* (suara keadaan di sebuah pameran), *restaurant* (suara keadaan di sebuah restoran), *station* (keadaan di stasiun kereta api), *street* (keadaan di jalan raya), *train* (keadaan didalam kereta api).

### 5.2 Ukuran Objektif Kualitas Sinyal Ucapan

Ukuran subjektif merupakan ukuran yang sesungguhnya untuk menentukan kualitas dan inteligibilitas sinyal suara. Akan tetapi, karena keterbatasan sumber daya, pengukuran subjektif tidak dapat dilakukan. Sebagai gantinya, ukuran objektif digunakan pada paper ini. Beberapa ukuran objektif yang memiliki korelasi tinggi dengan ukuran subjektif digunakan [25]. Pada paper ini, digunakan 3 ukuran objektif untuk mengukur kualitas dan inteligibilitas sinyal ucapan: Segmental SNR (SegSNR), Frequency Weighted SNR (FwSNR), dan perceptual evaluation of speech quality (PESQ). SegSNR umum dipakai untuk mengukur kualitas sinyal suara, akan tetapi korelasi antara SegSNR dengan ukuran subjektif kualitas sinyal ucapan rendah [25]. Oleh karena itu ukuran-ukuran lain juga dipakai sebagai pembanding. Pemilihan ukuran-ukuran ini dilakukan berdasarkan penelitian bahwa ukuran-ukuran ini memiliki korelasi yang tinggi dengan ukuran subjektif [25].

SegSNR mengikuti prinsip yang sama dengan SNR kecuali SegSNR merata-ratakan SNR untuk semua frame. SegSNR dihitung sebagai berikut [26]:





Gambar 3: Diagram blok untuk komputasi PESQ.

SegSNR =

$$\frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{n=Nm}^{Nm+N-1} x^2(t)}{\sum_{n=Nm}^{Nm+N-1} (x(t) - \hat{x}(t))^2} \right) \quad (29)$$

dimana  $x(t)$  adalah sinyal bersih,  $\hat{x}(t)$  adalah sinyal *enhanced*,  $N$  adalah panjang frame, dan  $M$  adalah jumlah total frame.

Sementara itu FwSNR adalah SegSNR yang dihitung dalam domain frekuensi. FwSNR dihitung menggunakan formula sebagai berikut [26]:

FwSNR =

$$\frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{k=1}^K B(m, k) \log_{10} \left[ \frac{X^2(m, k)}{(X(m, k) - \hat{X}(m, k))^2} \right]}{\sum_{k=1}^K B(m, k)} \right), \quad (30)$$

dimana  $B(m, k)$  adalah bobot yang diberikan pada band frekuensi ke- $k$ ,  $K$  adalah jumlah band frekuensi,  $X(m, k)$  adalah amplitudo sinyal bersih pada frekuensi ke- $k$ , dan  $\hat{X}(m, k)$  adalah amplitudo sinyal *enhanced*. Parameter  $B(m, k)$  dapat dihitung menggunakan formulasi sebagai berikut:

$$B(m, j) = |X(m, k)|^\gamma \quad (31)$$

dimana pada penelitian ini digunakan  $\gamma = 0.2$ .

PESQ adalah rekomendasi P.862 ITU-T pada tahun 2000 [27]. Struktur komputasi untuk PESQ

dapat dilihat pada Gambar 3 [28]. PESQ sendiri awalnya tidak digunakan untuk mengevaluasi kualitas sinyal ucapan pada *speech enhancement*. Akan tetapi, sejumlah penelitian menemukan bahwa PESQ memiliki korelasi sangat tinggi dengan hasil evaluasi menggunakan ukuran subjektif [25, 29]. Pada PESQ, digunakan nilai dari model sistem pendengaran manusia dalam membandingkan sinyal bersih dan sinyal *enhanced* untuk menentukan kualitas sinyal ucapan. PESQ dihitung menggunakan kombinasi linear dari rata-rata nilai gangguan  $D_{ind}$  dan nilai rata-rata gangguan asimetris  $A_{ind}$ . PESQ diformulasikan sebagai berikut [28]:

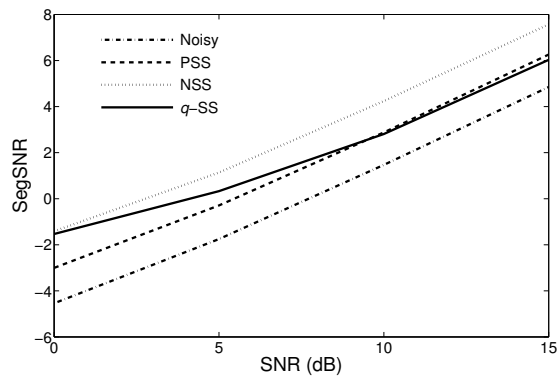
$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind}, \quad (32)$$

dimana  $a_0 = 4.5$ ,  $a_1 = -0.1$ , dan  $a_2 = 0.0309$ . Nilai-nilai ini, yang ditentukan secara empiris, memiliki korelasi optimum dengan ukuran kualitas sinyal suara secara subjektif [25]. Nilai PESQ berkisar antara 0.5 ke 4.5. Semakin tinggi nilai PESQ berarti semakin tinggi kualitas sinyal ucapan tersebut. Untuk lebih detail mengenai PESQ dapat mengacu ke [24, 27, 28].

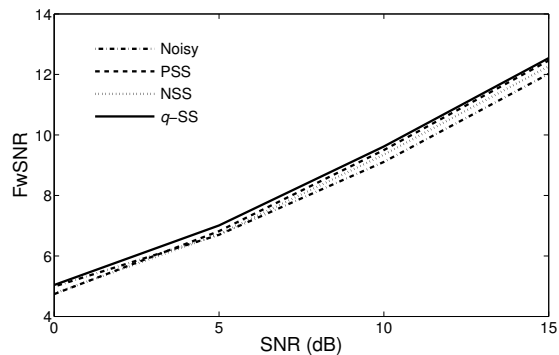
### 5.3 Evaluasi

Pada paper ini, akan dikomparasi performa tiga jenis metode *spectral subtraction*: PSS (Persamaan (5)), NSS (Persamaan (6)) dan  $q$ -SS (Persamaan (14)). Evaluasi menggunakan tiga ukuran objektif digunakan untuk ketiga metode ini. Untuk spektrum negatif, digunakan proses *flooring* dengan  $\beta = 0.002$  yang diterapkan untuk PSS, NSS, dan  $q$ -SS.

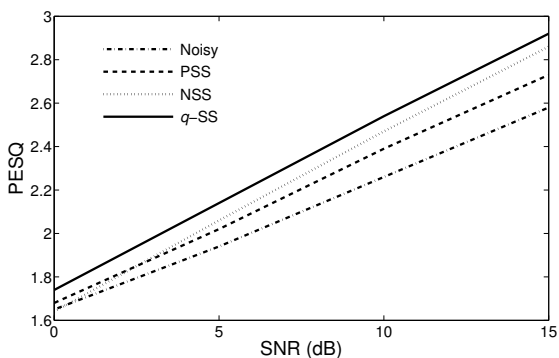




(a) Segmental SNR (SegSNR)



(b) Frequency weighted SNR (FwSNR)



(c) Perceptual Evaluation of Speech Quality (PESQ)

Gambar 4: Hasil evaluasi kualitas sinyal suara pada NOIZEUS database menggunakan tiga jenis metode spectral subtraction: PSS, NSS, dan  $q$ -SS. Digunakan tiga ukuran objektif: SegSNR, FwSNR, dan PESQ. Hasil ini diperoleh dengan merata-ratakan nilai tersebut untuk semua jenis derau pada database NOIZEUS.

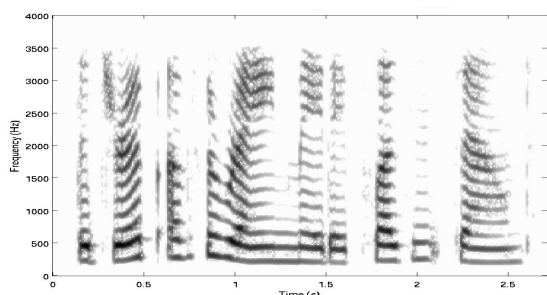
## 6. HASIL PERCOBAAN DAN DISKUSI

Gambar 4 menunjukkan perbandingan hasil evaluasi ketika metode PSS, NSS, dan  $q$ -SS menggunakan

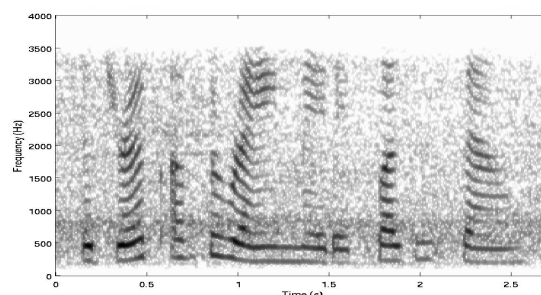
ukuran objektif SegSNR, FwSNR, dan PESQ. Hasil pada grafik adalah nilai rata-rata untuk kedelapan jenis derau dan untuk ketiga puluh kalimat pada NOIZEUS. Berdasarkan hasil ini dapat dilihat bahwa  $q$ -SS memberikan performa terbaik untuk ukuran FwSNR dan PESQ dibandingkan PSS dan NSS untuk setiap kondisi SNR. Sementara untuk SegSNR, NSS memberikan hasil terbaik. Secara intuitif, hasil ini mengindikasikan bahwa NSS memiliki tingkat *noise suppression* tertinggi. Namun, tingginya *noise suppression* tidak semerta-merta kualitas sinyal ucapan juga membaik, karena *noise suppression* yang tinggi juga dibarengi tingkat *speech distortion* yang juga tinggi. Selain itu, SegSNR juga memiliki korelasi yang rendah dengan kualitas sinyal suara yang sebenarnya. Kedua ukuran lain, FwSNR dan PESQ lebih baik dalam mengindikasikan kualitas sinyal ucapan. Oleh karena itu, hasil ini menunjukkan bahwa  $q$ -SS meningkatkan kualitas suara lebih baik daripada NSS dan PSS.

Hasil ini sudah dapat diduga berdasarkan kurva *Gain* kedua metode ini seperti dapat dilihat pada Gambar 1. Karena *flooring* terjadi pada SNR yang lebih rendah pada NSS, maka pada NSS, range SNR dimana derau akan dihilangkan lebih lebar dibandingkan  $q$ -SS. Akan tetapi, hal ini juga berakibat NSS berakibat distorsi yang lebih besar pada sinyal ucapan dibandingkan  $q$ -SS, terutama pada daerah SNR yang rendah. Dilain pihak,  $q$ -SS memiliki tingkat *suppression* derau yang lebih besar pada SNR tinggi dan  $q$ -SS dapat mencegah terjadinya distorsi sinyal ucapan ketika SNR rendah, dengan melakukan *flooring* pada SNR yang lebih tinggi. Namun ini berakibat menjadi lebih rendahnya tingkat *noise suppression*  $q$ -SS secara keseluruhan karena derau tidak dihilangkan pada kondisi SNR rendah. Kemampuan  $q$ -SS yang memiliki tingkat *noise suppression* lebih tinggi pada SNR yang tinggi namun *flooring* pada SNR lebih tinggi menjadi penting karena ketika SNR tinggi, sinyal ucapan mendominasi derau, sehingga estimasi derau dapat lebih baik dilakukan dan kemungkinan distorsi lebih kecil. Pada SNR rendah, derau lebih dominan pada sinyal, dan estimasi derau menjadi lebih sulit pada kondisi ini, sehingga peluang terjadinya distorsi justru lebih besar.

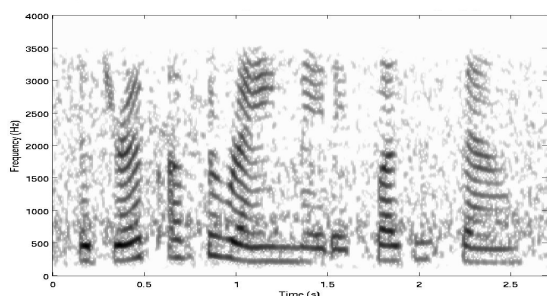
Gambar 5 adalah spektrogram sinyal ucapan yang bunyinya adalah: "The drip of the rain made a pleasant sound" untuk sinyal tanpa derau, dikontaminasi oleh derau suara mobil (car noise) pada SNR 10 dB, sinyal *enhanced* hasil PSS, sinyal *enhanced* hasil NSS, dan sinyal *enhanced* hasil  $q$ -SS. berdasarkan hasil spektrogram ini dapat kita amati derau musikal yang terjadi akibat PSS. Selain derau musikal, dapat dilihat juga masih banyak derau



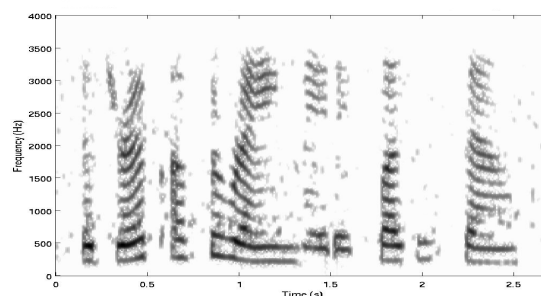
(a) Sinyal bersih (tanpa derau)



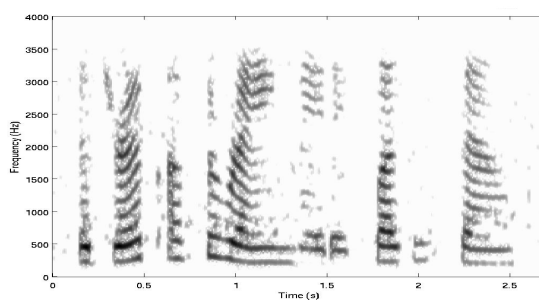
(b) Sinyal noisy



(c) Sinyal enhanced hasil PSS



(d) Sinyal enhanced hasil NSS

(e) Sinyal enhanced hasil  $q$ -SS

Gambar 5: Perbandingan spektrogram sinyal dengan kalimat: “The drip of the rain made a pleasant sound” keadaan bersih tanpa derau, ditambah car noise pada 10 dB SNR, hasil PSS, hasil NSS, dan hasil  $q$ -SS.

yang tertinggal setelah PSS diimplementasikan. Untuk spektrogram hasil NSS dan  $q$ -SS dapat dilihat jumlah derau yang tertinggal jauh berkurang dibandingkan PSS. Pada NSS dan  $q$ -SS, distorsi pada sinyal ucapan juga dapat diamati apabila dibandingkan dengan sinyal bersih. Banyak bagian dari sinyal ucapan yang hilang akibat NSS dan  $q$ -SS. Jika dibandingkan derau musikal pada NSS dan  $q$ -SS, pada keduanya masih ada derau musikal yang tertinggal. Ini ditunjukkan dengan masih adanya isolasi-isolasi puncak spektrum (berupa titik-titik hitam pada spektrogram). Jika diamati spektrogram kedua metode ini dapat dilihat bahwa jumlah dan tingkat ketebalan puncak yang terisolasi pada  $q$ -SS berkurang dibandingkan NSS. Perlu

dicatat, bahwa pada spektrogram, semakin tebal spektrogram mengindikasikan amplitudo yang juga semakin besar dan tingkat kenyaringannya juga lebih besar. Hal ini mengindikasikan bahwa  $q$ -SS mengurangi derau musikal lebih baik daripada NSS. Ketika hasil sinyal suara setelah  $q$ -SS didengarkan oleh penulis, derau musikal memang berkurang dibandingkan NSS.

## 7. KESIMPULAN DAN RENCANA KEDEPAN

Pada paper ini, telah ditunjukkan  $q$ -SS memiliki performa untuk meningkatkan kualitas sinyal suara menggunakan ukuran PESQ dan FwSNR lebih baik dibandingkan dengan menggunakan NSS dan PSS. Hal ini dikarenakan  $q$ -SS memiliki *Gain* lebih

rendah dibandingkan NSS pada SNR yang lebih tinggi sehingga dapat mengurangi derau lebih baik pada kondisi SNR tinggi. Selain itu pada  $q$ -SS, *flooring* juga terjadi pada SNR yang lebih tinggi dibandingkan NSS. Hal ini berakibat  $q$ -SS memiliki distorsi speech yang lebih kecil dibandingkan NSS pada kondisi SNR yang rendah tanpa mempengaruhi jumlah derau yang dihilangkan secara signifikan.

Penelitian lebih lanjut dibutuhkan untuk lebih meningkatkan performa  $q$ -SS. Selain itu, karena  $q$ -SS dapat diadaptasikan kepada berbagai varian NSS, maka akan menarik untuk menginvestigasi performa  $q$ -SS menggunakan formulasi NSS lainnya seperti menggunakan pendekatan multi-band, penggunaan model pendengaran dan lain-lain. Penelitian untuk menentukan  $q$  yang lebih baik berdasarkan properti sinyal suara seperti distribusi dan SNR juga menarik untuk diinvestigasi. Pengukuran kualitas dan inteligibilitas sinyal suara menggunakan ukuran subjektif juga perlu dilakukan untuk memperoleh hasil evaluasi yang lebih baik terhadap performa  $q$ -SS.

#### Daftar Pustaka

- [1] J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.
- [2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 27, no. 2, pp. 113–120, 1979.
- [3] W. Jin, X. Liu, M. Scordilis, and L. Han, "Speech enhancement using harmonic emphasis and adaptive comb filtering," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 2, pp. 356–368, 2010.
- [4] M. Sambur, "Adaptive noise canceling for speech signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 5, pp. 419–423, 1978.
- [5] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE*, vol. 80, no. 10, pp. 1526–1555, 1992.
- [6] R. Martin, "Statistical methods for the enhancement of noisy speech," pp. 43–65, 2003.
- [7] I. Cohen, "Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models," *Signal Processing*, vol. 86, p. 698709, 2006.
- [8] D. Farrokhi, R. Togneri, and A. Zaknich, "Speech enhancement of non-stationary noise based on controlled forward moving average," *Proc. Internat. Symp. Communications and Information Technologies*, pp. 1551–1555, 2007.
- [9] M. Gabrea, E. Grivel, and M. Najun, "A single microphone kalman filter-based noise canceller," *IEEE Signal Processing Letters*, vol. 6, no. 3, pp. 55–57, 1999.
- [10] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, pp. 208 – 211, apr 1979.
- [11] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (nss), hidden Markov models and the projection, for robust speech recognition in cars," *Speech Commun.*, vol. 11, no. 2-3, pp. 215 – 228, 1992.
- [12] R. M. Udrea, N. Vizireanu, S. Ciochina, and S. Halunga, "Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale," *Signal Processing*, vol. 88, no. 5, pp. 1299 – 1303, 2008.
- [13] S. Kamath and P. Loizou, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, pp. IV–4164, May 2002.
- [14] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *Speech and Audio Processing, IEEE Transactions on*, vol. 7, no. 2, pp. 126–137, 1999.
- [15] H. Pardede, K. Shinoda, and K. Iwano, "Q-Gaussian based spectral subtraction for robust speech recognition," *Proc. Interspeech*, pp. 1255–1258, 2012.
- [16] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans Acoust.*, vol. 28, no. 2, pp. 137 – 145, apr 1980.
- [17] V. Schless and F. Class, "Snr-dependent flooring and noise overestimation for joint application of spectral subtraction and model combination," *Proc. Internat. Conf. Spoken Language Process.*, 1998.
- [18] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Stat. Phys.*, vol. 52, pp. 479–487, 1988.
- [19] L. Nivanen, A. L. Méhauté, and Q. Wang, "Generalized algebra within a nonextensive statistics," *Rep. Math. Phys.*, vol. 52, no. 3, pp. 437 – 444, 2003.
- [20] C. Tsallis, "Entropic nonextensivity: A possible measure of complexity," *Chaos Solitons Fractals*, vol. 13, no. 3, pp. 371 – 391, 2002.
- [21] S. Furuichi, "Information theoretical properties of Tsallis entropies," *J. Math. Phys.*, vol. 47, no. 2, 2006.
- [22] S. Umarov, C. Tsallis, and S. Steinberg, "On a  $q$ -Central Limit Theorem consistent with nonextensive statistical mechanics," *Milan J. Math.*, vol. 75, pp. 307–328, 2008.
- [23] J. William A. Kretzschmar, "Speech as a complex system," in *The Linguistics of Speech*. Cambridge University Press, 2009.
- [24] P. Loizou, *Speech enhancement: theory and practice*, ser. Signal processing and communications. CRC Press, 2007.
- [25] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 16, no. 1, pp. 229–238, 2008.
- [26] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*, ser. Ellis Horwood Series in Artificial Intelligence. Prentice Hall PTR, 1988.

- [27] ITU, *Perceptual Evaluation of Speech Quality (PESQ) and Objective method for end to end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*. ITU-T Rec. P. 862, 2000.
- [28] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," *Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*, vol. 2, pp. 749–752 vol.2, 2001.
- [29] W. Liu, K. Jellyman, J. S. D. Mason, and N. Evans, "Assessment of objective quality measures for speech intelligibility estimation," *Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–I, 2006.