CHARACTERIZATION OF ACOUSTIC PARAMETERS OF RESONANT /m/ AND /n/ SOUNDS FOR THE DEVELOPMENT OF OBJECTIVE INDICATORS OF RESONANT VOICE THERAPY

SITI SUFIAH BINTI ABDUL RAZAK

UNIVERSITI SAINS MALAYSIA

2025

CHARACTERIZATION OF ACOUSTIC PARAMETERS OF RESONANT /m/ AND /n/ SOUNDS FOR THE DEVELOPMENT OF OBJECTIVE INDICATORS OF RESONANT VOICE THERAPY

by

SITI SUFIAH BINTI ABDUL RAZAK

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

October 2025

ACKNOWLEDGEMENT

Alhamdulillah, all praise and gratitude are due to Allah (SWT), the Most Merciful and Compassionate, for His countless blessings, guidance, and strength throughout this journey. By His grace, I was granted the perseverance and clarity to complete this research. Every insight and milestone is a testament to His mercy. My deepest thanks to my supervisor, Dr. Khairy Anuar bin Mohd Khairuddin, for his unwavering support, invaluable mentorship, and selfless dedication. His generosity in sacrificing his time and personal leisure to support my progress reflects a rare and admirable commitment to his students. His guidance and encouragement have been instrumental to both this thesis and my personal growth. I am deeply indebted to him and offer my sincerest thanks and appreciation. I am also sincerely grateful to my co-supervisor, Professor Dr. Mohd Normani bin Zakaria, for his constructive feedback and consistent support, which greatly enhanced the quality of this work. Special thanks to the faculty and staff of the School of Health Sciences (PPSK), Universiti Sains Malaysia Health Campus (USMKK), and the Institute of Postgraduate Studies (IPS) for their assistance throughout my postgraduate journey. I am forever indebted to my parents for their love, sacrifices, and unwavering belief in me. Their support has been my foundation. To my beloved spouse, Ahmad Nabil bin Aziz, thank you for your patience, encouragement, and constant support during the most challenging times. I pray that Allah accepts this humble effort, that any shortcomings be forgiven, and that this work brings benefit to others.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APQ11	11-point perturbation quotient
APQ3	3-point amplitude perturbation quotient
APQ5	5-point amplitude perturbation quotient
CPP	Cepstral peak prominence
CPPF0	CPP fundamental frequency
CPPFull	Non-segmented CPP
CPPSD	CPP standard deviation
CPPVoice	Voice-segmented CPP
DDA	Difference of differences of amplitude
DDP	Difference of differences of periods
EPQ	Extent perturbation quotient
F0	Fundamental frequency
F0r	Range for the fundamental frequency
F0SD	Standard deviation of fundamental frequency
F1	First formant
F2	Second formant
F3	Third formant
F4	Fourth formant
GNE	Glottal-to-noise excitation ratio
HNR	Harmonics-to-noise ratio
M	Mean
Malay CAPE-V	Malay consensus auditory-perceptual of voice
Min-Max	Minimum-maximum values
MSPEAK	Malay speech assessment kit

NHR	Noise-to-harmonic ratio
OME	Oral motor examination
PPQ	Period perturbation quotient
PPQ5	5-point perturbation quotient
RAP	Relative amplitude perturbation
RV	Resonant voice
RVT	Resonant voice therapy
SD	Standard deviation
SLT	Speech-language therapist
SPL	Sound pressure level
TV	Typical voice
VFEs	Vocal function exercises

LIST OF APPENDICES

PENCIRIAN PARAMETER AKUSTIK BUNYI RESONAN /M/ DAN /N/ UNTUK PENCIPTAAN PENUNJUK OBJEKTIF TERAPI SUARA RESONAN

ABSTRAK

Terapi suara resonan (RVT) ialah intervensi berasaskan bukti untuk gangguan suara, biasanya dimulakan dengan bunyi dengung seperti /m/ dan /n/. Suara resonan (RV) disahkan melalui maklum balas pendengaran dan sentuhan di bawah bimbingan ahli terapi pertuturan-bahasa (SLT), namun kaedah ini terhad di luar persekitaran klinikal. Kajian ini bertujuan mengenal pasti parameter akustik objektif yang boleh menunjukkan RV dan menyokong pembangunan alat maklum balas yang nyata bagi amalan kendiri. Seramai 90 peserta (45 perempuan, 45 lelaki) menyumbang data untuk /m/, dan 83 peserta (41 perempuan, 42 lelaki) untuk /n/. Selepas latihan RV, setiap peserta menghasilkan tiga rakaman 6 saat bagi RV dan suara tipikal (TV). Tiga saat bahagian tengah dianalisis menggunakan perisian Praat untuk mengekstrak frekuensi asas (F0), formant (F1–F4), sisihan piawai F0 (F0SD), ukuran gangguan, keunggulan puncak cepstral (CPPFull, CPPVoice), nisbah harmonik kepada hingar (HNR), dan intensiti purata. Perbezaan signifikan ($p < 0.05$, saiz kesan sederhana) ditemui pada APQ3 dan DDA bagi perempuan, dan intensiti purata bagi lelaki. Intensiti purata secara konsisten membezakan RV daripada TV, dengan nilai yang lebih tinggi dalam RV. Berbanding /n/, bunyi /m/ mempunyai corak ukuran gangguan lain dan F0SD yang lebih rendah, manakala CPP dan HNR mempunyai corak lebih tinggi dalam RV. Analisis formant menunjukkan perbezaan keputusan signifikan pada F2 dan F3 dalam lelaki, manakala perempuan kurang konsisten. Penemuan ini menunjukkan potensi parameter akustik membezakan RV daripada TV dan menyokong penggunaannya sebagai penunjuk objektif bagi pembangunan alat maklum balas RVT.

CHARACTERIZATION OF ACOUSTIC PARAMETERS OF RESONANT /m/ AND /n/ SOUNDS FOR THE DEVELOPMENT OF OBJECTIVE INDICATORS OF RESONANT VOICE THERAPY

ABSTRACT

Resonant voice therapy (RVT) is an evidence-based intervention for voice disorders, often initiated with nasal sounds such as /m/ and /n/. Resonant voice (RV) is usually verified through auditory and tactile feedback under a speech-language therapist's (SLT) guidance, but these perceptual methods are limited outside clinical settings. The present study aimed to identify objective acoustic parameters that could indicate RV and support tangible feedback tools for self-guided practice. Ninety participants (45 females, 45 males) contributed data for /m/, and 83 (41 females, 42 males) for /n/. After RV training, each provided three 6-second recordings of RV and typical voice (TV). The middle 3 seconds were analyzed in Praat to extract fundamental frequency (F0), formants (F1–F4), F0 standard deviation (F0SD), perturbation measures, cepstral peak prominence (CPPFull, CPPVoice), harmonic-to-noise ratio (HNR), and mean intensity. Significant differences ($p < 0.05$, medium effect size) were found in APQ3 and DDA for females, as well as in mean intensity for males. Mean intensity consistently distinguished RV from TV, with significantly higher values in RV. Compared to the /n/ sound, the /m/ sound presented with lower value perturbation measures and F0SD and higher value for CPP and HNR, though not significant. Formant analyses revealed significant F2 and F3 differences in males, while female results were less consistent. These findings demonstrate the potential of acoustic parameters to differentiate RV from TV. Consistent patterns across multiple measures, some statistically supported, highlight their promise as objective indicators and their utility in developing tangible feedback tools to enhance RVT practice.

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Like other behavioral intervention approaches, voice therapy involves teaching patients techniques to develop new vocal behaviors while simultaneously facilitating the replacement of old, well-established, and often maladaptive vocal habits (Iwarsson, 2014). Voice therapy requires patients to practice the learned techniques both during and outside of therapy sessions to promote positive treatment outcomes (van Leer and Connor, 2012). During therapy sessions, patients benefit from the feedback and guidance of the speech-language therapist (SLT) to ensure the accuracy of practice. However, the same close monitoring is not available outside therapy sessions. Patients have often reported difficulty in accurately applying the learned techniques without SLT expert feedback (van Leer and Connor, 2010), which may affect the effectiveness of the prescribed voice treatment. Therefore, the availability of tangible feedback to facilitate the mastery of the learned techniques may be beneficial for patients to engage in self-guided practice outside of clinical sessions.

1.2 Resonant Voice Therapy

Resonant voice therapy (RVT) is one of the techniques that is effective in treating voice disorders (Yiu et al., 2017). This technique aims to generate voice more easily and more resonantly (Roy et al., 2003) with the least effort and impact on the vocal folds (Ramig and Verdolini, 1998; Verdolini et al., 1998). This type of voice that is central to RVT is termed resonant voice (RV). In comparison, a typical voice (TV) refers to the habitual or baseline manner of phonation without intentional resonance enhancement. Fundamentally, RV follows the same basic physiological process as TV production, involving vibration of the vocal folds initiated by airflow from the lungs. This vibration is then modified by the resonating cavities of the vocal tract to produce the desired acoustic output. In TV, resonance is diffusely distributed across the oral and pharyngeal cavities, whereas in RV it is directed upward and forward into the midfacial area (Smith et al., 2005; Titze, 2001). TV also typically involves balanced adduction with efficient closure of the vocal folds, while RV uses relatively neutral, non-overadducted postures during phonation (Verdolini et al., 1998; Verdolini-Marston et al., 1995). In addition, during the production of RV, the laryngeal vestibule was judged to be narrower compared to that of TV (Ufema and Montequin, 2001; Yanagisawa et al., 1989).

These alterations in the vocal tract configuration, optimized vocal fold posture, and focused resonance yield a clearer, more radiated sound, accompanied by stronger vibratory sensation in the midfacial region, particularly at the alveolar ridge and maxillary bones (Chen et al., 2014) while minimizing inter-vocal fold impact stress (Roy et al., 2003; Verdolini-Marston et al., 1995). These features make RV both efficient in maximizing vocal output and protective in reducing the risk of vocal fold injury.

At the basic training gesture, the acquisition of RV is usually facilitated by utilizing nasal sounds such as /m/ or /n/ as stimuli (Stemple et al., 2018; Verdolini Abbott, 2008). As nasal sounds, /m/ and /n/ are produced by lowering the soft palate to allow the sounds to be channeled through the nasal cavity and finally to the nasal orifices, i.e., around the midfacial area. These configurations of /m/ and /n/ sound productions make them ideal for eliciting a more forward sound characterizing RV.

Once this basic level is established, the complexity of the tasks is increased to words, phrases, sentences, and conversations, with a combination of other vowels and consonants, while maintaining RV (Stemple et al., 2018; Theis and Carlson, 2022; Verdolini Abbott, 2008) until generalization into daily communication is achieved. At any level, the patients make use of their auditory feedback, i.e., a perception of a strong and clear sound, and tactile feedback, i.e., a sensation of increased midfacial vibration, as indicators of the attainment of RV (Yiu et al., 2012).

1.3 Characteristics of Acoustic Parameters of Resonant Voice

Acoustic parameters are quantitative measures derived from the analysis of acoustic signals produced by the vocal tract (Hillenbrand, 2011). Given their measurable nature, these parameters provide a reliable and objective means of characterizing the voice (Dejonckere et al., 2001; Patel et al., 2018). Previous studies have investigated a range of acoustic parameters to characterize RV in comparison to TV, to identify objective indicators of RV (Andrade et al., 2013; Aydınlı et al., 2019; Barrichelo and Behlau, 2007; Barrichelo-Lindström and Behlau, 2009; Ogawa et al., 2014; Smith et al., 2005; van Leer et al., 2015; Vlot et al., 2016).

Several of these parameters have been shown to distinguish RV from TV, supporting their use as objective indicators of RV. These parameters span a range of acoustic domains, including frequency-related measures (e.g., fundamental frequency [F0], F0 range [F0r]), perturbation indices (e.g., period perturbation quotient [PPQ], amplitude perturbation quotient [APQ], extent perturbation quotient [EPQ]), spectral characteristics (e.g., formants i.e., first formant [F1], second formant [F2], third formant [F3], and fourth formant [F4], spectral differences), cepstral measures (e.g., cepstral peak prominence [CPP], CPP fundamental frequency [CPPF0], CPP standard deviation [CPPSD]), and noise-related metrics (e.g., glottal-to-noise ratio [GNE]). Their consistent application across literature underscores the relevance of acoustic parameters in characterizing RV and supports their potential integration into RVT practice.

1.4 Problem Statement

The training process in RVT typically begins with patients producing RV on single sounds, most commonly /m/ and /n/, before progressing to more complex speech tasks that require sustained use of the acquired RV. During therapy sessions, an SLT is able to provide expert feedback that supports the accurate production of these resonant targets. However, outside the clinical setting, the absence of SLT guidance and the inherently subjective nature of perceptual judgments on RV attainment may affect the consistency and accuracy of its production.

To address this issue, the use of tangible feedback to determine the attainment of RV may offer a promising solution for self-guided practice among patients. In this regard, previous studies have demonstrated that RV is associated with various acoustic parameters. Owing to their numerical nature, these acoustic parameters can be transformed into analytical dimensions, serving as objective indicators that are instrumental in the development of tangible feedback for RV attainment.

Despite growing interest in the use of acoustic parameters to characterize RV, the number of empirical studies remains limited. Moreover, each available study has employed a relatively narrow set of acoustic parameters. As a result, there is currently no consensus on which parameters are most appropriate or reliable for identifying RV, particularly in sustained productions of /m/ and /n/ at the basic training level of RVT. This lack of consistency in parameter selection, combined with the limited scope of existing research, underscores the need for further investigation to identify a representative set of acoustic parameters that can support the objective identification of RV. To begin addressing this gap, the present study explores a broader set of acoustic parameters in sustained /m/ and /n/ sounds, to contribute to a more comprehensive understanding of the acoustic characteristics of RV that are relevant to the development of objective indicators for RV attainment.

It is important to note that to meaningfully characterize the acoustic parameters of RV, it is necessary to establish a comparison with TV productions, as practiced by previous studies. This comparison enables the identification of specific acoustic parameters that differentiate RV from TV. By analyzing the same sounds, i.e., /m/ and /n/, in both RV and TV conditions, it becomes possible to determine with greater clarity which parameters reflect changes attributable to RV.

1.5 Research Questions

a) What are the characteristics of the acoustic parameters of resonant /m/ and /n/ sounds?

b) What are the characteristics of the acoustic parameters of typical /m/ and /n/ sounds?

c) Are there any differences in the acoustic parameters between resonant and typical /m/ and /n/ sounds?

1.6 Study Objectives

1.6.1 General objective

To identify potential objective indicators of resonant voice, the present study aims to characterize the acoustic parameters of resonant /m/ and /n/ sounds by measuring and comparing them with their respective typical counterparts.

1.6.2 Specific objectives

a) To measure each acoustic parameter characterizing the typical /m/ and /n/ sounds.

b) To measure each acoustic parameter characterizing the resonant /m/ and /n/ sounds.

c) To compare each acoustic parameter of the resonant /m/ and /n/ sounds to their respective typical /m/ and /n/ sounds.

1.7 Study Hypothesis

1.7.1 Null hypothesis

There are no statistically significant differences in any of the acoustic parameters between resonant and typical productions of /m/ and /n/ sounds.

1.7.2 Alternate hypothesis

There are statistically significant differences in one or more acoustic parameters between resonant and typical productions of /m/ and /n/ sounds.

1.8 Significance of the Study

The ultimate goal of the present study is to support the development of an application (app) that facilitates the practice of RVT. The acoustic findings on resonant /m/ and /n/ sounds will serve as a foundation for the app's design, beginning with the verification of RV attainment at the isolation level, i.e., the basic training gestures. Establishing reliable acoustic indicators at this level is a critical step, as these data can later serve as reference points for identifying RV in more complex speech tasks, such as words, phrases, sentences, and conversation. In the long term, the app is envisioned as a digital tool that provides tangible feedback and structured guidance to help patients acquire and maintain RV outside clinical sessions, ultimately contributing to more consistent practice and improved RVT outcomes.

CHAPTER 2

LITERATURE REVIEW

2.1 Voice Therapy

Voice therapy is a therapeutic intervention that aims to change the behaviour associated with voicing to promote effective phonation (Gartner-Schmidt et al., 2013). Many voice therapy techniques have been introduced to treat voice disorders, with most of them successfully reducing voice deficiency and increasing phonation efficiency (Chen et al., 2007). In general, voice therapy techniques can be grouped into physiologic voice therapy and symptomatic voice therapy (American Speech-Language-Hearing Association, n.d.-a). Physiologic voice therapy focuses on improving the function of the vocal mechanism by addressing respiration, phonation, and resonance aspects of voice production (Stemple et al., 2014). Examples of physiologic voice therapy include the RVT, accent method, and vocal function exercises (VFEs) (American Speech-Language-Hearing Association, n.d.-a). Meanwhile, symptomatic voice therapy aims to improve specific vocal symptoms, such as roughness, breathiness, and altered pitch and loudness (Stemple et al., 2014). Examples of symptomatic voice therapy techniques include confidential voice therapy, inhalation phonation, twang therapy, and the yawn-sigh technique (American Speech-Language-Hearing Association, n.d.-a).

2.2 Resonant Voice Therapy

Evidence supports the effectiveness of RVT as a therapeutic approach for voice disorders. A systematic review by Yiu et al. (2017) reported that existing evidence supports the use of RVT for treating voice disorders and concluded that there is sufficient evidence for its effectiveness. Although there are several variations of RVT in the literature, such as Lessac-Madsen resonant voice therapy (Verdolini Abbott, 2008), humming (Raymond et al., 2006), Lessac's Y-Buzz (Barrichelo and Behlau, 2007), and resonance therapy (Stemple et al., 2018), each of them adopted the use of RV, characterized by a clear and strong sound accompanied by increased midfacial vibration, beginning with single sounds and progressing to generalization at the daily conversational level.

2.3 Acoustic Studies on Resonant Voice Attainment

Reflecting clinical practice, most previous studies have relied on subjective judgment to determine the attainment of RV, particularly based on patients' and SLTs' auditory feedback and patients' tactile feedback (i.e., midfacial vibration) (Yiu et al., 2017). Although limited, there are studies that have explored the potential of using objective parameters as indicators to facilitate the attainment of RV. Among these, a total of eight studies utilized acoustic parameters as objective measures to determine the attainment of RV (i.e., Andrade et al., 2014; Aydınlı et al., 2019; Barrichelo & Behlau, 2007; Barrichelo-Lindström & Behlau, 2009; Ogawa et al., 2014; Smith et al., 2005; van Leer et al., 2015; Vlot et al., 2017).

From these acoustic-based studies, the sample sizes ranged from 6 to 98 participants, with females (138 participants) outnumbering males (127 participants). The studies covered a broad age range from 16 to 84 years. Normophonic participants were recruited in five studies (Andrade et al., 2013; Aydınlı et al., 2019; Barrichelo and Behlau, 2007; Barrichelo-Lindström and Behlau, 2009; Smith et al., 2005), with two focusing on typical voice users (Andrade et al., 2013; Aydınlı et al., 2019), another two on professional voice users (e.g., actors and singers) (Barrichelo and Behlau, 2007;

Barrichelo-Lindström and Behlau, 2009), and one including both typical and professional normophonic voice users (Smith et al., 2005). The incorporation of both normophonic and dysphonic typical voice users was explored in two other studies (Ogawa et al., 2014; Vlot et al., 2016). The least represented group of voice users is dysphonic professional voice users, who were examined in only one study (van Leer et al., 2015). A summary of these participant demographics is presented in Table 2.1.

Table 2.1 Participants' Demography

Studies	**Number**	**Male**	**Female**	**Age**	**Types of Participants**
Smith et al. (2005)	6	3	3	Range: 23-43	Normophonic: Typical and professional voice users
Barrichelo and Behlau (2007)	9	6	3	Range: 16-28	Normophonic: Professional voice users
Barrichelo-Lindström and Behlau (2009)	54	23	31	Mean male:23 Mean female: 24	Normophonic: Professional voice users
Andrade et al. (2013)	23	7	16	Not available	Normophonic: Typical voice users
Ogawa et al. (2014)	20	15	5	Range: 24-72	Normophonic: Typical voice users
	21	16	5	Range: 35-84	Dysphonic: Typical voice users
van Leer et al. (2015)	14	2	12	Range: 16-72	Dysphonic: Professional voice users
Vlot et al. (2016)	49	29	20	Range: 24-71	Normophonic: Typical voice users
	49	16	33	Range: 20-82	Dysphonic: Typical voice users
Aydınlı et al. (2019)	20	10	10	Range: 20-30	Normophonic: Typical voice users

Table 2.2 summarizes the characteristics of RV training, presenting the techniques used to train and elicit RV, as well as the training schedule. In terms of RV techniques, seven studies employed a single technique. In relation to this, humming was the most frequently used, appearing in four studies (Andrade et al., 2013; Aydınlı et al., 2019; Ogawa et al., 2014; Vlot et al., 2016), followed by the Lessac Y-Buzz technique in two studies (Barrichelo-Lindström and Behlau, 2009), and Lessac-Madsen resonant voice therapy (van Leer et al., 2015) in one study. Finally, only one

study utilized either Boone and McFarlane's forward focus technique or Lessac Y-Buzz (Smith et al., 2005).

Regarding the training schedule, most studies did not provide any information about the practice required for participants to achieve RV, except for three studies. Among these, the duration of training sessions typically ranged from 15 to 30 minutes. For the number of sessions, one study delivered the training in a single session (Aydınlı et al., 2019), while one study distributed the training across four sessions, either weekly or every other day within a week (Barrichelo-Lindström and Behlau, 2009). One study did not explicitly outline the complete training schedule, but reported that the training occurred after at least two preceding voice therapy sessions (van Leer et al., 2015).

Table 2.2 Characteristics of Resonant Voice Training

Studies	**Technique**	**Schedule**
Smith et al. (2005)	Boone and McFarlane, forward focus, or Lessac Y-buzz	Not available
Barrichelo and Behlau (2007)	Lessac Y-buzz	Not available
Barrichelo-Lindström and Behlau (2009)	Lessac Y-buzz	4 sessions of 30 minutes each, once a week
Andrade et al. (2013)	Humming	Not available
Ogawa et al. (2014)	Humming	Not available
van Leer et al. (2015)	Lessac Madsen resonant voice therapy	At least 2 preceding voice therapy sessions
Vlot et al. (2016)	Humming	Not available
Aydınlı et al. (2019)	Humming	1 session of 15-30 minutes

A summary of the methods used to determine RV attainment, including the verification procedures by external raters, is presented in Table 2.3. Three studies employed a combination of auditory feedback from researchers or SLTs and tactile feedback reported by participants (Aydınlı et al., 2019; Smith et al., 2005; Vlot et al., 2016). Another three studies relied solely on auditory feedback (Andrade et al., 2013; Barrichelo-Lindström and Behlau, 2009; van Leer et al., 2015). In two of these, the feedback was self-reported by participants (Barrichelo-Lindström and Behlau, 2009; van Leer et al., 2015), while in one study, it was determined by the SLT (Andrade et al., 2013). The remaining two studies used tactile feedback alone, based on participants' own perceptions, to confirm the attainment of RV (Barrichelo and Behlau, 2007; Ogawa et al., 2014).

Findings on external verifications show that it was incorporated in half of the studies. Of the four studies that incorporated external verification (Aydınlı et al., 2019; Barrichelo and Behlau, 2007; Barrichelo-Lindström and Behlau, 2009; Smith et al., 2005), the number of raters ranged from two to four. Regarding raters' characteristics, two studies employed SLTs trained in voice (Barrichelo and Behlau, 2007; Barrichelo-Lindström and Behlau, 2009). Another study also used SLTs as raters, but their voice training was not specified (Aydınlı et al., 2019). One additional study included raters with experience in voice quality evaluation, though their professional backgrounds were not reported (Smith et al., 2005).

Table 2.3 Methods to Determine the Attainment of Resonant Voice

Studies	**Subjective Indicator**	**External Verification**
Smith et al. (2005)	Tactile-participant Auditory-researcher	Yes 3 judges experienced in voice quality rating
Barrichelo and Behlau (2007)	Tactile-participant	Yes 3 SLTs trained in voice
Barrichelo-Lindström and Behlau (2009)	Auditory-participant	Yes 4 SLTs trained in voice
Andrade et al. (2013)	Auditory-SLT	No
Ogawa et al. (2014)	Tactile-participant	No
van Leer et al. (2015)	Auditory-participant	No
Vlot et al. (2016)	Tactile-participant Auditory-researcher	No
Aydınlı et al. (2019)	Tactile-participant Auditory-SLT	Yes 2 SLTs

Table 2.4 outlines the types of tools and stimuli used for analyses in RV and TV. In terms of analysis tools, most acoustic parameters were assessed using widely available voice analysis software, with Praat being the most frequently used, appearing

in four studies (Andrade et al., 2013; Barrichelo-Lindström and Behlau, 2009; Ogawa et al., 2014; Vlot et al., 2016). Notably, one study utilized a custom-developed iOS mobile application for its analysis (van Leer et al., 2015).

There were various stimuli used in the production of RV and TV. Several studies used the same stimuli in RV and TV productions. These stimuli were the production of single sounds /a/ (van Leer et al., 2015), /i/ (Barrichelo and Behlau, 2007; Barrichelo-Lindström and Behlau, 2009; Smith et al., 2005), and /m/ (Aydınlı et al., 2019), repetition of single syllable /me/ (van Leer et al., 2015), and the production of nasal and non-nasal sentences (Aydınlı et al., 2019). Some studies compared different stimuli for RV and TV, specifically comparing the vowel sound /a/ with the nasal sound /m/ (Andrade et al., 2013) or the vowel sound /e/ with the nasal sound /m/ (Ogawa et al., 2014; Vlot et al., 2016).

Table 2.4 Types of Tools and Stimuli Used for Analyses

Studies	Tools	Analyzed Stimulus	
		RV	TV
Smith et al. (2005)	Windaq	A prolonged /i/ in /θriː/ during the counting of “1,2,3”	A prolonged /i/ in /θriː/ during the counting of “1,2,3”
Barrichelo and Behlau (2007)	VoxMetria	A prolonged /i/ in /ji/	A prolonged single /i/
Barrichelo-Lindström and Behlau (2009)	Praat	A prolonged /i/ in /ji/	A prolonged single /i/
Andrade et al. (2013)	Praat	A prolonged /m:/ in humming	A prolonged single /a/
Ogawa et al. (2014)	Praat	A prolonged /m:/ in humming	A prolonged single /e:/
van Leer et al. (2015)	iOS-based CPP application	A prolonged single /a/ Repetition of /me/	A prolonged single /a/ Repetition of /me/
Vlot et al. (2016)	Praat	A prolonged /m:/ in humming	A prolonged single /e:/
Aydınlı et al. (2019)	Analysis of Dysphonia in Speech and Voice (ADSV) software	A prolonged /m:/ in humming Non-nasal sentence, /vʌlı vʌlızını vɛɾdı/ Nasal sentence, /mʌjmuŋ mʌmʌdʌŋ mɛmŋuŋdu/	A prolonged /m:/ in humming Non-nasal sentence, /vʌlı vʌlızını vɛɾdı/ Nasal sentence, /mʌjmuŋ mʌmʌdʌŋ mɛmŋuŋdu/

2.3.1 Acoustic comparison between resonant voice and typical voice

Table 2.5 summarizes the acoustic parameters investigated in previous studies comparing RV and TV, aimed at characterizing RV production. Among the parameters, F0 was utilized in four studies (Andrade et al., 2013; Barrichelo-Lindström and Behlau, 2009; Ogawa et al., 2014; Smith et al., 2005). Although the differences were not statistically significant, these studies consistently found that F0 was higher in RV than in TV, except for female participants in Barrichelo-Lindström and Behlau (2009), where similar F0 values were observed in both voice types. In a related investigation of first formant tuning, Smith et al. (2005) also found that both F0 among females and the second harmonic (H2) among males were higher in RV compared to TV. In addition to F0, Andrade et al. (2013) also examined fundamental frequency range (F0r), a derivative measure of F0, and reported lower values in RV compared to TV. However, none of these differences reached statistical significance.

Various perturbation measures were employed in previous studies (i.e., Barrichelo & Behlau, 2007; Ogawa et al., 2014; Vlot et al., 2017). For period perturbation measures, the parameters included were EPQ (Barrichelo and Behlau, 2007), PPQ (Barrichelo and Behlau, 2007), and PPQ5 (Ogawa et al., 2014; Vlot et al., 2016). Except for PPQ5 among the normophonic population in the study by Vlot et al. (2016), the other examined parameters consistently showed lower values in RV compared to TV, across gender and types of participants. Notably, significant differences were found for EPQ among normophonic participants (Barrichelo and Behlau, 2007) and for PPQ5 among dysphonic participants (Ogawa et al., 2014). For amplitude perturbation measures, only one type of parameter, i.e., APQ11, was examined across two studies (Ogawa et al., 2014; Vlot et al., 2016). Although these studies included both normophonic and dysphonic populations, the APQ11 findings were reported without separating results by gender, instead presenting them as a combined group. Generally, APQ11 showed lower values in RV compared to TV, except in the normophonic population examined by Vlot et al. (2016), where RV values were higher compared to TV. Additionally, this parameter consistently showed statistically significant differences in the dysphonic groups across both studies.

In addition to the aforementioned perturbation measures that assessed vocal stability, another stability-related parameter, i.e., irregularity, was examined in one study (Barrichelo and Behlau, 2007). The findings indicated that irregularity was not only lower in RV compared to TV, but the difference was also statistically significant.

Spectral measures were examined in three studies (i.e., Andrade et al., 2014; Barrichelo-Lindström & Behlau, 2009; Smith et al., 2005). Specifically, the parameters included were formants (F1 to F4) (Barrichelo-Lindström and Behlau, 2009), and spectral differences such as between the first formant and fundamental frequency (F1–F0) (Andrade et al., 2013; Barrichelo-Lindström and Behlau, 2009) and between the first formant and second harmonic (F1–H2) (Barrichelo-Lindström and Behlau, 2009). The study that utilized formants reported a consistent trend across both males and females, with lower values in RV compared to TV. In fact, the differences in formants were statistically significant, except for F1 among males. As for the spectral differences, although the F1–F0 values did not reach statistical significance, the findings showed lower values in RV compared to TV for both genders, consistent across studies (Andrade et al., 2013; Barrichelo-Lindström and Behlau, 2009). A more substantive finding was observed in F1–H2, which showed a statistically significant difference, with lower values in RV compared to TV among males (Barrichelo-Lindström and Behlau, 2009).

Investigation using CPP was reported in two studies (Aydınlı et al., 2019; van Leer et al., 2015). Across both studies, the findings consistently showed higher CPP values in RV compared to TV across stimuli. It is important to note that, although CPP values from a prolonged /m/ sound were planned to be analysed in Aydınlı et al.'s (2019) study, these findings were not reported in the results. In addition to CPP, variations of CPP, such as CPPF0 and CPPSD, were also investigated in Aydınlı et al. (2019)'s study. The findings related to CPPF0 were particularly noteworthy, as they revealed a significant result, with RV showing higher values than TV (Aydınlı et al., 2019). For CPPSD, the findings consistently showed lower values in RV compared to TV, although the difference was not statistically significant (Aydınlı et al., 2019).

Noise-related measures were investigated in only one study (i.e., Smith et al., 2005). These noise-related measures included GNE and a noise parameter. The findings revealed that GNE was higher in RV compared to TV. Meanwhile, the noise parameter was lower in RV than in TV. However, neither finding was statistically significant.

Table 2.5 Comparison of Acoustic Parameters Between Resonant Voice and Typical Voice Across Studies

Category	Acoustic Parameters		Study
F0	F0r: RV < TV		Andrade et al. (2013)
	F0: RV > TV		
	Normophonic	Dysphonic	
	F0: RV > TV	F0: RV > TV	Ogawa et al. (2014)
	Female	Male	
	F0: RV = TV	F0: RV > TV	Barrichelo-Lindström and Behlau (2009)
	F0: RV >TV	2F0: RV >TV	Smith et. al (2005)
Perturbation Measures	EPQ: RV < TV*		Barrichelo and Behlau (2007)
	PPQ: RV < TV		
	Normophonic	Dysphonic	
	PPQ5: RV < TV	PPQ5: RV < TV*	Ogawa et al. (2014)
	PPQ5: RV > TV	PPQ5: RV < TV	Vlot et al. (2016)
	APQ11: RV < TV	APQ11: RV < TV*	Ogawa et al. (2014)
	APQ11: RV > TV	APQ11: RV < TV*	Vlot et al. (2016)
Irregularity	RV < TV*		Barrichelo and Behlau (2007)
Spectral measures	Female	Male	Barrichelo-Lindström and Behlau (2009)
	F0: RV = TV	F0: RV > TV	
	F1: RV < TV*	F1: RV < TV	
	F2: RV < TV*	F2: RV < TV*	
	F3: RV < TV*	F3: RV < TV*	
	F4: RV < TV*	F4: RV < TV*	
		F1-H2: RV < TV*	
	F1-F0: RV < TV*	F1-F0: RV < TV	
	F1-F0: RV < TV	F1-F0: RV < TV	Andrade et al. (2013)
CPP	A prolonged single /a/		van Leer et al. (2015)
	CPP: RV > TV*		
	Repetition of /me/		
	CPP: RV > TV*		
	Prolonged /m:/ in humming		Aydınlı et al. (2019)
	CPP: Not available		
	Nasal sentence		
	CPP: RV > TV		
	CPPSD: RV < TV		
	Non-nasal sentence		
	CPP: RV > TV*		
	CPPSD: RV < TV		
	CPPF0: RV > TV*		
Noise	GNE: RV > TV		Barrichelo and Behlau (2007)
	Noise: RV < TV		

Note:
RV -Resonant voice
TV -Typical voice
* -Significantly different

2.4 Acoustic Parameters

Instead of relying on a single acoustic parameter, the present study adopts multiple acoustic measures to objectively characterize RV. This comprehensive approach enables more robust and accurate differentiation between RV and TV, thereby enhancing the precision of RV identification. The acoustic parameters to be utilized include F0, formants (i.e., F1 to F4), F0 standard deviation (F0SD), amplitude perturbation measures (i.e., shimmer, 3-point APQ [APQ3], 5-point APQ [APQ5], 11-point APQ [APQ11], and difference of differences of amplitudes [DDA]), period perturbation measures (i.e., jitter, relative amplitude perturbation [RAP], 5-point PPQ [PPQ5], and difference of differences of period [DDP]), CPP (i.e., CPPFull and CPPVoice), harmonic-to-noise ratio (HNR), and mean intensity. A detailed description of each acoustic parameter is presented below.

1. F0

   This parameter is defined as the lowest frequency of a periodic waveform (Gerhard, 2003). It is an acoustic measure that reflects the vibrating rate of vocal folds during voice (Lee and Humes, 2012) and is measured in hertz (Hz). Additionally, it is correlated with the perception of habitual pitch. An individual's habitual pitch is influenced by their gender and age (Eichhorn et al., 2018). In general, prepubescent children, regardless of whether boys or girls, have the highest F0 of about 300 Hz (Hunter, 2009), followed by adult females with an F0 of about 211 Hz, and the lowest F0 is found among adult males, approximately 110 Hz (Re et al., 2012).

2. Formants

   The resonant frequencies of the air in the vocal tract during speech production are known as formants (Dissen et al., 2019; Ladefoged and Johnson, 2015).

Given that different formants are associated with different vocal tract configurations (Aalto et al., 2018; Ladefoged and Johnson, 2015), speech sounds produced using different configurations will result in different formant characteristics.

3. F0SD

Variability in F0 over time reflects the stability and consistency of the voice and is quantified using the F0SD parameter (Gelin Li et al., 2021). Specifically, the parameter measures the variability of F0 over time, reflecting pitch stability during speech (Bowen et al., 2013). A lower value of F0SD suggests more stable voice characteristics, whereas a high F0SD indicates greater pitch variability and is typically observed in voices with less stability (Nguyen et al., 2024).

4. Perturbation

This parameter refers to variability or irregularity within a system (Titze, 1994). When applied to voice, it describes the cycle-to-cycle variability in the vocal signal (Boon et al., 2013). Two commonly used perturbation measures are amplitude perturbation measures and period perturbation measures, which are described below.

a. Amplitude perturbation measures

These parameters quantify variations in the amplitude of a voice signal across successive vocal cycles (Heiberger and Horii, 1982). There are various parameters in amplitude perturbation measures. Although varied, these parameters are interpreted in the same way, i.e., lower values are observed in better voice characteristics, and higher values are observed in poorer voice characteristics (Leclerc et al., 2013). In the present study, only shimmer, variation of amplitude perturbation quotients (APQ) (i.e., APQ3, APQ5, APQ11), and DDA are used. These parameters are described below.

i. Shimmer

Variations in the amplitude of the sound wave or the intensity of vocal emission are known as shimmer (Wertzner et al., 2005). It is typically measured as a percentage (Teixeira and Fernandes, 2014).

ii. Amplitude perturbation quotient

Among the amplitude perturbation quotients, APQ3 uses two neighboring periods for comparison, APQ5 uses a wider window of four neighboring periods, and APQ11 extends this analysis by averaging amplitude variations across eleven successive periods to assess longer-term amplitude instability (Wu et al., 2017). Despite the differences in the number of periods considered, all three parameters measure the same underlying concept of amplitude perturbation over varying temporal windows. They are expressed in percentage (%).

iii. DDA

The average absolute second-order differences in amplitude values across consecutive cycles provide a measure of amplitude fluctuation, known as DDA (Wu et al., 2017). This parameter is also measured in percentage (%).

b. Period perturbation measures

These parameters measure the cycle-to-cycle variation in the fundamental frequency of the vocal signal (Teixeira and Fernandes, 2014; Teixeira et al., 2013). Although there are multiple ways to assess it, the pattern remains consistent, i.e., lower values are associated with better voice characteristics, while higher values indicate poorer voice characteristics (Cortés Ponce et al., 2024; Rachel et al., 2017; Upadhya et al., 2017). The period perturbation measures used in the present study are jitter, RAP, PPQ5, and DDP. These parameters are described below.

i. Jitter

The variability or perturbation of the fundamental frequency is known as jitter (Behlau et al., 2001; Wertzner et al., 2005). This measure is typically expressed as a percentage (Teixeira and Fernandes, 2014).

ii. RAP

It calculates how much a period differs from the average of itself and its two adjacent periods, normalized by the mean period length (Teixeira and Fernandes, 2014). The value of this parameter is a percentage (%).

iii. PPQ5

This variation reflects the average absolute variation between a single period and the mean of that period and its four nearest neighbors, divided by the overall average period (Teixeira and Fernandes, 2014). This parameter is expressed as a percentage (%).

iv. DDP

It represents the average of the absolute differences between successive jitter cycles, giving insight into variability between changes (Wu et al., 2017). This parameter is measured in percentage (%).

5. CPP

Derived from the cepstrum of a sound wave, this acoustic parameter measures the relative amplitude of the peak corresponding to the dominant harmonic in the voice signal (Patel et al., 2018; Watts et al., 2017). The present study will focus on using CPPFull and CPPVoice. For CPPFull, the CPP is calculated across the entire frequency range of the signal, including both harmonic components from voiced sounds and nonharmonic components from unvoiced sounds (Boersma and Weenink, 2023). Meanwhile, CPPVoice focuses specifically on voiced segments and captures the more prominent cepstral peak that arises from the harmonic structure of voiced speech (Boersma and Weenink, 2023). For both CPPFull and CPPVoice, better voice characteristic is typically associated with high CPP values, vice versa (Antonetti et al., 2020). This parameter is quantified in terms of decibels (dB) (Antonetti et al., 2020; Aydınlı et al., 2019).

6. HNR

The parameter measures the ratio of harmonic (periodic) energy to noise (aperiodic) energy in the voice signal (Fernandes et al., 2018). Higher HNR values indicate more periodic and stable vocal fold vibration, while lower HNR values reflect greater noise and irregularity in the signal (Fernandes et al., 2018). This pattern suggests that higher HNR will be observed in better voice characteristics, vice versa. As it is a ratio, this measure is dimensionless and does not carry any unit.

7. Mean intensity

This parameter represents the average loudness or energy (Titze and Sundberg, 1992), typically measured in dB. Intensity naturally increases due to specific vocal tract adjustments, particularly pharyngeal widening and laryngeal vestibule widening (Smith et al., 2005). These adjustments create a more efficient acoustic system, allowing for higher intensity with less effort, which is associated with better voice characteristics.

The above acoustic parameters were selected based on current recommendations by Patel et al. (2018), as they are considered among the most reliable for characterizing voice in both clinical and research settings. These include F0, F0SD, and CPP, which are particularly effective in capturing vocal pitch characteristics and spectral regularity. The inclusion of these parameters ensures alignment with current best practices in acoustic voice analysis and contributes to the robustness of the study's methodology.

In addition to the parameters suggested by Patel et al. (2018), the present study also includes acoustic measures that are widely available in Praat (Boersma and Weenink, 2023), which continues to be a standard tool for voice analysis. These include formants (F1 to F4), period perturbation measures, i.e., jitter, RAP, PPQ5, and DDP, and amplitude perturbation measures, i.e., APQ3, APQ5, APQ11, and DDA. These perturbation measures are sensitive to the voice signal and are particularly useful for identifying subtle differences in phonatory control. Additional parameters such as HNR and mean intensity are also included to capture the periodicity and loudness of the voice signal.

2.4.1 Directional trends of acoustic parameters

It is important to note that certain acoustic parameters exhibit a well-established directional trend based on theoretical knowledge, whereby lower or higher values are associated with specific voice characteristics. However, some parameters do not show