Language in Society (2025), 1–26
doi:10.1017/S0047404525000053

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The 'Critical Role' of voice quality in Dungeons and Dragons: A case study of non-player characters voiced by Matthew Mercer

Zac Boyd and Míša Hejná

Aarhus University, Denmark
Address for correspondence: Míša Hejná; Jens Chr. Skous Vej 4, 8000, Aarhus C, Denmark. Email: misa.hejna@cc.au.dk

Abstract

The current study provides a holistic analysis of voice quality and how it is employed via affective stancetaking through high performance of non-player characters on *Critical Role,* a popular Dungeons and Dragons digital media 'actual-play' series. Specifically, we ask how a character's moral stance is indexed through improvised performed speech. We show that current acoustic methods for voice quality have the potential for underrepresentation of sociolinguistically meaningful variation when relying solely on acoustic data. By incorporating both acoustic and auditory data, we find that constricted laryngeal settings (and whisperiness in particular) are used to signal evilness and negative moral stance, while unconstricted laryngeal settings (breathiness in particular) are employed to signal friendliness and positive moral stance. The two general vocal settings show nuanced variation linked to affective stancetaking, including one-off changes in characters' stances as well as their habitual styles. (Voice quality, stance, methods, affect, morality)*

Introduction

'High performance' (Coupland 2007) is a valuable resource for examining indices as it necessarily draws on existing language ideologies. In this article we explore how one may index aspects of morality in high performance. Morality and notions of good and evil have been claimed to be unique to human existence (Kagan 2000; Ayala 2010). Yet despite the fact that it is something fundamental to who we are and how we view and interact with the world, morality is nevertheless not particularly targeted by sociolinguists (for obvious reasons) in comparison to other social factors such as age, gender, social class, and so on (though see Shoaps 2009, Lippi-Green 2012, and attitudinal work in forensic linguistics). However, by looking at improvised high performance of relative morality performed via character personae, we may shed light on

overarching indexical relationships which may be broadly associated with notions of good and evil character archetypes. One such aspect that may be employed to this end is voice quality. Similarly to relative morality, voice quality remains relatively understudied within the sociolinguistic literature (though see the section *Voice quality* below), although voice quality is often seen as highly meaningful paralinguistically (Laver 1980/2009). In the present article we examine a process of affective stancetaking through voice quality variation to index relative morality in high performance role-playing interaction. We note that anytime we discuss performative speech throughout this article, it should be taken to be understood as 'high performance' (Coupland 2007), as opposed to everyday performance of speech as a discursive construction of social identity and personae widely studied in sociolinguistic literature.

We do so through an intraspeaker analysis of one Dungeon Master (DM, see the next section) for Dungeons and Dragons (D&D), a fantasy narrative tabletop role-playing game (TTRPG). Data for the present article comes from voice actor Matthew Mercer, the DM for *Critical Role* (*CR*), a popular digital media D&D series. We ask how Mercer employs aspects of voice quality to convey facets of good and evil character morality in Non-Player Characters' (NPCs) identities (in other words, characters who inhabit the world of the role-playing experience and who are not controlled or voiced by players at the table) in a performative setting. Specifically, what (if any) are the consistencies across characters within this vocal production based on if the character is an Ally or Enemy (or 'characterological figures' (Agha 2005) who draw on semiotic representations of good and evil)?

We follow Pratt (2023), Starr, Wang, & Go (2020), and Yanushevskaya, Gobl, & Ní Chasaide (2018), who call for affect to be theorised into (socio)linguistics. Adopting concepts from the fields such as psychology and media studies is by no means novel in sociolinguistics (e.g. Accommodation Theory; Lawson & Giles 1973), but there are still a number of psychological aspects of human existence that remain unexplored by linguists. While Starr and colleagues (2020) focus specifically on sensuality, here we target performed morality. Though claims about inner character (and by extension, morality) are notoriously difficult to assess and quantify (for obvious reasons), exploring phonatory variation through the lens of a well-documented D&D series allows for such an examination. Given that we know the motivations of the characters and the context in which it is presented, it is clear what moralities are being represented via Mercer's characterisation.

With this in mind, the present article explores how Mercer employs extreme phonatory variation to signal aspects of performed morality (i.e. good and evil) through affective stancetaking in relation to characters' relationship to the players being an Ally, an Enemy, or Neutrally Aligned to the group. Our results provide evidence that Mercer employs similar aspects of phonatory variation in affective stancetaking based on performed personae's alignment to the adventuring party (the in-game players' characters) and that character's overarching moral leanings. Furthermore, such performances (specifically of evilness) are similarly portrayed regardless of if the character is human (or human-esque) or if that character is a vocal representation of a non-human entity.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

Theoretical framework

Dungeons and Dragons and Critical Role

D&D is a role-playing game in which players and the DM collaboratively construct a long-form narrative story driven by the actions of the players through gameplay. In this, the DM serves as storyteller, referee, and orchestrator of the game, guiding the players through the narrative and managing the game-world's challenges and outcomes. First introduced in 1974, D&D has seen a massive resurgence in popularity since the late 2010s, resulting in a sort of renaissance period of the game. The discourse within the D&D community attributes part of this massive resurgence in D&D's popularity to the success of *CR*.

CR is an actual-play[1] D&D program which is broadcast on Twitch streaming service, YouTube digital video platform, and is released via podcast. The cast of *CR* comprises eight "nerdy-ass voice actors" (as Mercer states during the intro to each episode). Mercer himself is a well-established American voice actor whose primary acting experience is in anime voiceover dubbing and video games. He is also the DM for *CR*'s main campaigns, meaning he is not only the arbiter of the game rules but is also the one who builds the world and gives life and voice to the NPCs.

While a core aim of this article is a discussion of how Mercer signals character morality (i.e. good versus evil), the concepts of 'good' and 'evil' have specific meanings and implications in D&D that differ from other settings. 'Good' and 'evil' are directly related to fundamental roleplaying elements and the cosmology/lore of the D&D universe, for example, a creature's 'Alignment'. So as to not conflate discussions of character morality with the structural alignment of D&D, we rely on an Ally/Enemy distinction as a proxy for examining 'good' and 'evil'. This systematic distinction is particularly important for readers who are familiar with the alignment system present in D&D, and we note that structural alignment is not considered at any point within our analysis. Importantly, the rest of this article refers to 'Alignment' in broader strokes, specifically in terms of a character's alignment TOWARDS THE ADVENTURING PARTY being either Aligned as an Ally, an Enemy, or being Neutral to the players. For our analysis this Ally/Enemy distinction results in a near one-to-one mapping to morality judgements of 'good' and 'evil' as would typically be viewed both within and outside D&D.

Playing the role: Personae and stancetaking in 'high performance'

The speech setting discussed in the present article is complex and relatively unique (though can be found in other actual-play broadcasts)—it is both performative yet conversational, containing (at times intimate) interactional situations which are simultaneously being viewed by thousands of people. Furthermore, this type of speech deviates from the typical type of speech data used in sociolinguistic research more broadly in that much of the speech produced by Mercer is intended to represent speech from non-human entities. It is the performance itself that indexes the social meanings discussed within.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

We state this purely to acknowledge that all sounds created here are 'human sounds' in that they are being produced by a human, but the ideological representation may be that of an interpretation of both human (or human-like) creatures as well as non-human-ness.

Waskul & Lust (2004:336) suggest that TTRPGs share great similarities with improvisational theatre and such a comparison is even more apt for *CR*. Indeed, Mercer IS performing, but his performance is first and foremost aimed at the players at the table, with the wider audience being generally peripheral to this performance. Although the speech context is conversational, Mercer as the DM actively guides the conversation. He switches on the fly between countless characters, and in the context of D&D the characters represent something akin to numerous PERSONAE, and the performances of these personae (or characters) employ variation which is constructed to set the stage (so to speak) for Mercer and the players to more readily engage with the wider narrative.

In the present data Mercer draws on characterisations which enact 'strategic inauthenticity' (Coupland 2007) to create a performance which is 'indexing a social identity AND the fact that it is not [his] own' (Bell & Gibson 2011:564, original emphasis; see also Mackay 2001:156). In doing so, Mercer is able to portray characters which draw on socially recognizable roles that can be enacted and performed through semiotic expression—what Agha (2005, 2011) calls 'figures of personhood'—as he manifests affective stances which display characters' moral leanings. The reflexivity of high performance reveals stance as implicit to performed speech (Jaffe 2009:11–12) as performers draw on 'characterological figures' (Agha 2005) which are tied to recognizable cultural forms.

Affective stancetaking relates directly to the emotional state of the speaker (Kiesling 2018). These stances also entail speaker intent, and an understanding of the emotional stakes and a character's intent is key for players to know how to respond to situations presented within the context of D&D. As Jaffe states, stancetaking provides a medium for speakers to 'attribute intentionality, affect, knowledge… and lay claim to particular… moral identities' (2009:9). If Mercer adopts a particular stance which the players attribute to be good or evil, it not only provides a narrative pathway for the story to develop, but those stances which position Mercer's character as good or evil, as an Ally or Enemy, allow the players to make appropriate choices in positioning THEMSELVES in alignment with or against that character. These affective stances further allow for NPCs to attempt to strategically manipulate and deceive the players by exploiting expected norms which draw on indexical representations of good and evil (see below).

As argued in Pratt (2023) and Podesva & Callier (2015), affective stancetaking is a useful analytic tool for examining voice quality, particularly in this setting. Indeed, Pratt (2023:20) argues that incorporation of affect in sociolinguistics 'represent[s] an approach to sociolinguistic style that sees affect as social and relational, and as a crucial piece of how speakers position themselves in their worlds'. The performative context of D&D allows us to examine how various personae convey their moral standing through these affective stances in a controlled environment by exploring phonatory variation within a single individual.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

Voice quality

Voice quality has been defined in many different ways (Esling, Moisik, Benner, & Crevier-Buchman 2019). We define voice quality as variation in sounds produced primarily in the larynx. Voice quality is important in the implementation of phonological contrasts as well as in signalling social meaning. Nevertheless, due to the methodological complexities and challenges surrounding vocal variation (e.g. Starr 2015; Garellek 2022), it has not attracted as much attention of sociolinguists as other speech related phenomena. Yet, work on emotions and attractiveness often conducted in the fields of speech engineering, evolutionary and social psychology (e.g. Hill & Puts 2021), as well as some of the sociolinguistic work available, strongly suggests that voice quality is essential for our understanding of spoken language. As Garellek (2022:18) puts it, 'a theory of spoken language needs a theory of the voice', and social factors form an important part of the whole vocal picture. In the field of sociolinguistics, voice quality has nonetheless received fairly limited attention (when compared to other aspects of sociolinguistic work) but has recently seen growing interest.

The methodology surrounding voice quality has been challenging because it has been recognised as highly multidimensional, both in production and perception. Phonologically oriented research frequently focuses on one to three phonation types employed for phonological contrasts (e.g. Garellek 2020:131–32, 2022). By contrast, research available on vocal variation in evolutionary and social psychology tends to focus primarily on fundamental frequency (F0) and pitch (Pisanski & Feinberg 2018). The diverse methodological approaches are understandable, but we cannot assume that measures that are able to capture the limited categorical differences in phonation that are required for the study of phonological contrasts will be sufficient to describe the wider range of voice quality differences that carry social meaning.

The work on the sociolinguistics of voice quality nonetheless provides us with a range of approaches. First, we find work falling within all three waves of sociolinguistic research (Eckert 2012), with a focus on macro-categories such as gender and ethnicity (e.g. Stuart-Smith 1999; Coadou 2006), and more recently also third-wave approaches with a focus on styles, personae, and stances (e.g. Podesva 2007; Mendoza-Denton 2011; Moisik 2012; Esposito 2016; Hejná & Eaton 2025). Second-wave studies have focused on overall vocal profile assessments in specific communities. This has been usually done with the use of audiovisual analyses relying on listening to recordings while visually inspecting waveforms and spectrograms in acoustic software such as Praat (Boersma & Weening 2023); or via auditory analyses, such as the Vocal Profile Analysis protocol (e.g. San Segundo, Foulkes, French, Harrison, Hughes, & Kavanagh 2018). Third-wave studies typically focus on one to three aspects of voice quality, such as the presence of creak, breathiness, and falsetto, and the variation found in their duration, with a focus on the construction of specific personae. In some recent voice studies, affect and affective stancetaking is taken into account very explicitly. Notably, Pratt (2023) shows how cisgender male students position themselves through stylistic choices,

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

particularly the 'chill' affect, associated with anti-institutional stances and artistic commitment. The use of linguistic and bodily signs, such as creak, speech rate, and posture, contributes to the ideological distinctions and reflects the students' responses to the school's academic and artistic provisions within the broader societal context of 'chill'. Furthermore, Starr (2015) finds that the presence or absence of 'sweet voice' is used to construct different types of female characters in Japanese anime cartoons, combined with features of Japanese women's language. More specifically, spoiled princesses as comedic villains are constructed with different vocal profiles and linguistic features than the sympathetic female characters who use sweet voice.

Methods

Character selection and character profiles

Data for the present study comes from voice actor Matthew Mercer, the DM for *CR*, specifically looking at NPC characters voiced by Mercer. Our analysis focuses on *CR*'s *Campaign 2: The Mighty Nein.*[2] Campaign 2 (one can think of this as 'season 2' or the cast's second adventure featuring new characters), which takes place within a high fantasy genre typical of classic D&D, featuring a mediaeval-like setting with magic, supernatural creatures, and fictional non-human races. Campaign 2 aired from January 11, 2018 until June 3, 2021, containing 141 episodes (average runtime 3h56m) accumulating a total of 483.6 hours of gameplay audio.[3]

In total, Mercer introduced 1,144 unique NPC characters during *The Mighty Nein's* campaign (see n. 3). Given the sheer number of characters, it was not feasible for us to examine each voice, especially considering the vast majority of these characters had little impact on the story beyond world-building and basic game mechanic functionality (i.e. a guard blocking a necessary entrance, etc.). While we are undoubtedly missing out on interesting and socially meaningful linguistic information by the exclusion of many of these characters, the core aim of this article is an in-depth analysis of character personae as it relates to wider notions of performed moral stances. In this, we do not suggest that these minor characters do not draw on aspects of personae, but rather that their character motivations are not necessarily as established as the more complex characters. This makes an in-depth character analysis problematic. As such, our criterion for NPC inclusion within this study relied on the following considerations: (i) the NPCs discussed within were all deemed to be central to the plot of the campaign story arcs (i.e. core plot points cannot have occurred without the character's involvement); (ii) the characters need to have been discussed by the adventuring party without them being present (in turn informing their importance as being central to the plot); and (iii) they have over ten minutes of collated audio data (with one exception, Uk'otoa;[4] see Table 1). We selected nineteen characters who met such criteria. In total, our dataset includes twelve male[5] characters, six female characters, and one non-/multi-gendered 'hive-mind'. Summary data for each character can be found in Table 1. The 'hive-mind', Somnovem, comprises four unique voiced

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

personalities, though for the purposes of analysis (except where explicitly noted below) we treat them as one character. All instances of overlapping speech between Mercer and the player characters, as well as speech from three live shows, was excluded from analysis.

For each character, we compiled a character profile, which includes all known information regarding the character, to be used in conjunction with that character's vocal profile. The character profiles consisted of any known demographic information, such as relative age, gender, occupation, race/

Table 1. List of NPC characters examined.

Name	Total audio	Alignment	Gender	Race/Creature type[6]
Artagan (The Traveller)	53:10	Ally	male	archfey
Astrid Becke	10:45	Ally	female	human
Avantika	18:28	Ally → Enemy	female	half-elf
Babenon Dosal (The Gentleman)	53:37	Ally	male	(water) genasi
Essek Thelyss	1:53:54	Neutral → Ally	male	drow
Isharnai, the Prism Sage	12:14	Enemy	female	hag
Lucien	45:16	Neutral → Enemy	male	tiefling
Ludinus Da'leth	20:03	Neutral	male	elf
Marion Lavorre	35:45	Ally	female	tiefling
Obann	10:50	Enemy	male	cambion
Orly Skiffback	20:46	Ally	male	tortle
Pumat Sol	47:55	Ally	male	firbolg
The Somnovem (The Hive-mind)	15:23	Enemy	non-/ multi-gender	unspecified aberration
Trent Ikithon	22:25	Enemy	male	human
Uk'otoa	02:26	Enemy	male	leviathan
Vess DeRogna	36:31	Neutral	female	half-elf
Vilya (Viridian)	41:07	Ally	female	half-elf
Yeza Brenatto	14:40	Ally	male	halfling
Yussa Errenis	33:34	Neutral → Ally	male	elf

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

creature type, and class (if available—meaning official class type of character within the D&D framework, e.g., wizard, bard, etc.). Furthermore, these profiles included detailed descriptions of the information known regarding the character's backstory and campaign story arc. Through this information we were able to make informed judgements regarding characters' relative morality and alignment to the players.

Vocal profiles

Character profiles were integrated with vocal profiles, involving a holistic analysis of Mercer's voice quality across each character. Vocal profiles were crafted through audiovisual analysis based on spectrograms, independent of the character profile analysis. The vocal profiles were created by one person knowledgeable of D&D and *CR* (first author), and a novel listener with no prior associations to either (second author). Initially, the vocal profiles were established through in-depth auditory and visual spectrogram analysis conducted by the novel listener, later corroborated by both authors for each character. Table 1 outlines the amount of speech available per character. The audiovisual analysis considered various aspects of voice quality, including mean pitch, pitch dynamism, modal voice, breathiness, whisperiness, creakiness, falsetto, tremor, and speech tempo, based on the Laryngeal Articulator Model framework (Esling et al. 2019). The following sections provide definitions of vocal characteristics and elaborate on the acoustic measures used in the second stage of the vocal profile analysis.

An example of modal voice is shown in Figure 1. What characterises canonical modal voice is a clear presence of formant structures and glottal pulses visible across a range of frequencies, identifiable as well-defined striations in the spectrogram (see Esling et al. 2019:44–47). Breathy voice corresponds to a percept of a softer voice quality, associated with lower intensity due to an increased airflow coming through the glottis and the larynx (Esling et al. 2019:56–58). Acoustically, the formant structure is less pronounced (compare Figure 1 and Figure 2), showing more friction in higher frequency energies.

Breathiness and whisperiness can and have been conceptualised as two points on the same acoustic and articulatory continuum, with whisperiness involving more friction and more airflow (Moisik, Hejná, & Esling 2019). However, as shown by Moisik and colleagues (2019) and Esling and colleagues

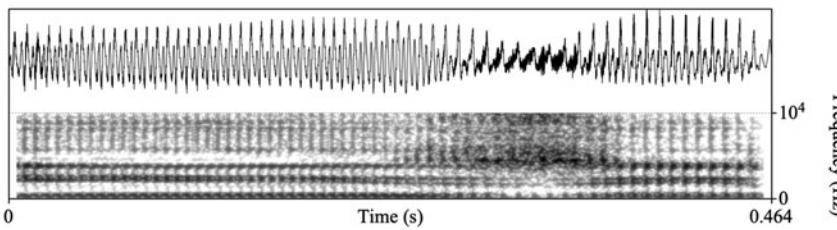

Figure 1. Modal voice.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

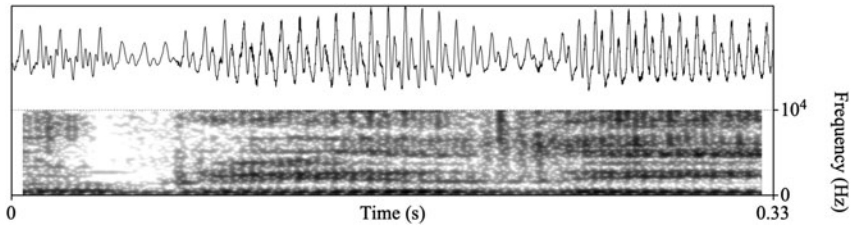

Figure 2. Breathy voice.

(2019:53–60), whisperiness is associated with a laryngeal constriction above the vocal folds, that is, in the epilarynx. Perceptually, whisperiness is noisier (higher-intensity) than breathiness (also Laver 1980/2009:45, 115). Figure 3 shows an example of whisperiness typically found in our data. Both breathiness and whisperiness can involve the presence or absence of vocal fold vibration. In our dataset, there is a tendency for the former to be voiced and for the latter to be voiceless. In Figure 3 we see the presence of energies at specific frequencies with relatively narrow bandwidths, which marks the presence of whisperiness. In some characters, the epilaryngeal constriction used to produce whisperiness includes vibration of the aryepiglottic folds, which results in growl, or harsh voice (see Figure 4).

Our analysis of creak was marked as present if the signal contained aperiodic vocal fold vibration and/or a sudden drop in F0. However, creak was rarely found in our data. While Mercer often does employ creak and generally has a more relaxed affect while narrating as DM, this is not the case relative to when he initiates an NPC voice. The avoidance of creak in character performances may be a means of distinguishing his non-stylised narrative DM style from the more exaggerated NPC performance styles. However, an analysis of Mercer's narrative DM style is beyond the scope of the current study. Similarly, while our analyses also included the identification of falsetto, this vocal profile was extremely rare. We therefore do not comment on falsetto in the rest of the article, and creak is only touched upon.

For acoustic analysis, each character had at least two pseudo-randomly selected two-minute speech segments, covering both 'Neutral' and 'non-Neutral' Emotional Contexts when possible. The selection was based on plot points, with no consideration for specific vocal qualities or content. However, obtaining both contexts was not possible for all characters, including one Neutrally Aligned (Ludinus Da'leth) and three Enemy Aligned characters (Somnovem, Isharnai, and Uk'otoa). The Neutral Emotional Context typically consists of informative gameplay elements such as lore building, giving relevant information, and quest related dialogue. Non-Neutral Emotional Contexts contain more emotionally charged speech. These are moments in which Mercer plays to the character's motivations, goals, and specifically their ties and relationships to members of the adventuring party. In doing so, these moments are contextually more emotionally loaded situations, conveying love, concern, anger, maliciousness, and so on, and these emotions

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

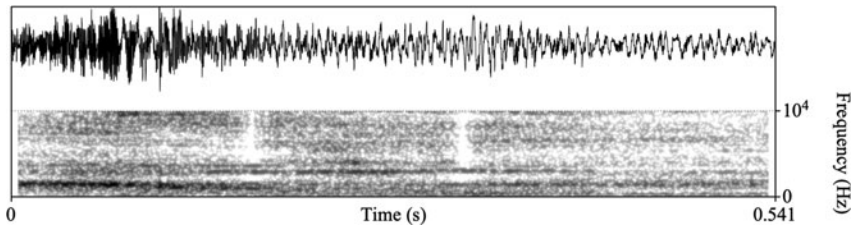

Figure 3. Whisper (voiceless whisperiness).

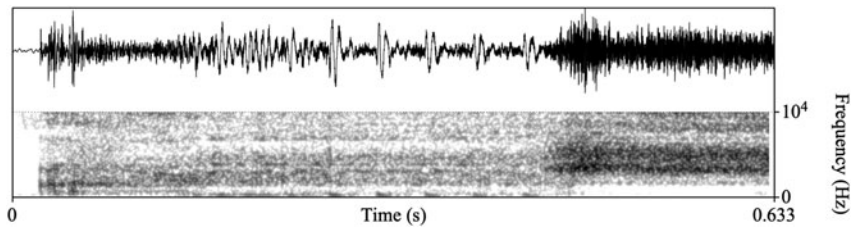

Figure 4. Whisper with growl.

are often directly correlated with Alignment and moral stancetaking. It makes sense that Enemy characters are not going to show love or concern for people who are actively trying to stop them from achieving their goals. Similarly, an Allied character is unlikely to threaten the adventuring party. Given this, we should expect to see different phonatory profiles for the different Emotional Contexts depending on the Alignment of a character. At the same time, where this variation in Emotional Contexts is available, it enables us to see whether a more/less evil stance is correlated with a specific phonation. In what follows, we show that this non-Neutral Emotional Context tends to be the locus of sociolinguistically relevant variation within our dataset (see the next section).

Data was transcribed in ELAN (2022) and automatically aligned with the Montreal Forced Aligner (McAuliffe, Socolof, Stengel-Eskin, Mihuc, Wagner, & Sonderegger 2022). VoiceSauce (Shue, Keating, Vicenik, & Yu 2011) was used to extract the measurements of interest in all vowels present in the selected speech samples. Measurements were taken at three intervals of each vowel and only the middle interval, corresponding to the second third of the vowel, was used for further analyses. The measures selected for extraction included F0, Cepstral Peak Prominence (CPP), H1*-H2*, H1*-A1*, and Harmonics-to-Noise Ratio < 3500 Hz (HNR35). The asterisks indicate normalisation (see Shue et al. 2011 for more details on these measures and the process of normalisation).

CPP and HNR35 fall in the category of noise measures of voice quality. They quantify the amount of noise in the signal. Specifically, CPP (measured in dB)

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

presents a spectrum of a spectrum (a cepstrum, or the inverse Fourier transform of a spectrum; Murton, Hillman, & Mehta 2020) and measures the dominance of the first rahmonic and its relationship to the overall noise in the signal (Fraile & Godino-Llorente 2014). It has repeatedly emerged as a robust measure of dysphonia severity (Fraile & Godino-Llorente 2014; Murton, Hillman, & Mehta 2020) and has been used primarily to quantify differences between modal and breathy voice, and modal and creaky voice. Lower values are associated with increased breathiness and creakiness (e.g. Kelterer & Schuppler 2019; Hejná, Šturm, Tylečková, & Bořil 2020), and increased dysphonia and voicelessness (Murton, Hillman, & Mehta 2020).

HNR35 stands for Harmonics-to-Noise-Ratio in the frequency range of 0–3500 Hz. The less harmonic and periodic a signal is, the lower the values of this measure. Garellek (2020:149–52) shows that creaky, modal, and breathy-creaky vowels in !Xóõ (Taa) show higher HNR35 values than pharyngealised vowels, which in turn show higher HNR35 values than breathy vowels, and harsh vowels are associated with the lowest HNR35 values. H1*-H2* and H1*-A1* belong to the family of voice quality measures based on spectral tilt, with the former representing the difference in the amplitude of the first and the second harmonics (measured in dB) and the latter that in the intensity of the first harmonic and the amplitude of the first formant (measured in dB). Higher values have been repeatedly observed to correspond to more lax phonatory settings (such as breathiness) and low values to more constricted phonatory settings (such as creakiness) (e.g. Gordon & Ladefoged 2001; Keating, Esposito, Garellek, Khan, & Kuang 2011). Similar to other measures based on the spectral slope, H1*-H2* and H1*-A1* rely on the presence of periodicity in the signal (Murton, Hillamn, & Mehta 2020:1596). While H1*-H2* has been shown to be a fairly reliable acoustic correlate of breathy and modal voice in particular (Keating et al. 2011), spectral tilt measures are dependent on reliant F0 tracking and require further data processing to increase their accuracy (Garellek 2020:143).

Other aspects of vocal variation were analysed qualitatively. These included pitch tremor, egressive as opposed to ingressive airstream mechanisms used to produce speech, and pitch dynamism. Tremor is linked with abrupt irregularities in the production of pitch. Ingressive sounds are produced while inhaling (rather than exhaling, as is the case with egressive sounds), and pitch dynamism refers to the variability in pitch.

Analysis

Voice quality as a stance of 'good' or 'evil'

Both the audiovisual and acoustic analyses revealed several patterns in how characters employ breathiness and whisperiness which correlate directly with their Alignment towards the players. Allies share overall voice qualities that range from modal to breathy, and also reveal a higher pitch dynamism. Enemies have a tendency to be portrayed with whisperiness and a limited pitch dynamism, and more individually can also display tremor and constricted

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

supralaryngeal settings, as well as atypical pausing patterns or speaking rates. The Neutral characters examined are almost always modal, occasionally drawing on breathiness.

We argue that whisperiness and breathiness are employed as a form of performed stancetaking. In terms of Enemies or evil characters, they may take stances of general 'evilness' wherein they employ intense whispery phonation throughout their entire speaking time, essentially signalling 'Here I am. I am evil'. This can be explained by a combination of the frequency code (Ohala 1995) and the links between emotions and laryngeal settings (and indeed overall physical states of speakers). Larger laryngeal structures can lead to increased irregularities, and in this sense, the irregular aspects of whisper likely signal a larger size, and indirectly also dominance and aggression (see Puts, Gaulin, & Verdolini 2006 and Anikin, Pisanski, Massenet, & Reby 2021 on these connections). At the same time, breathiness is associated with a lax laryngeal setting, while whisperiness is achieved by the opposite (Moisik et al. 2019). It should therefore not be surprising that phonatory profiles produced with tenser laryngeal constellations would be associated with evil characters. However, this only happens when the characters are saliently and OVERTLY evil (i.e. Isharnai, Uk'otoa, and the Somnovem). In this, race/creature type in our dataset is often (though not always) tied directly to Alignment and Morality, with Enemy characters often being non-human creatures who are typically associated with evilness (e.g. hag, leviathan, cambion, other Aberrations, etc.). Evil characters with more complex and nuanced personalities (aka, those whose social motivations require manipulation or cooperation to achieve their goals) will employ whisperiness in more emotionally charged situations and when conveying their true goals and motivations. This meshes well with our explanation above, because concealing one's moral disposition is a key aspect of manipulative behaviour. In this, evil characters draw on prototypical notions of exaggerated and pathologically affected voices, and/or voices of individuals portrayed as deformed (Kjeldgaard-Christiansen, Hejná, Clasen, & Eaton 2023), to index non-human-like threats and 'evilness', and whisperiness is the primary resource employed to that end, similarly to overtly evil characters. This is in line with Kjeldgaard-Christiansen and colleagues (2023), who employ analyses of *The Exorcist* to suggest that pathological and pathologically aged voices are employed to portray evil characters (the demon Pazuzu in this case, who claims to be 'the Devil himself') because of existing metaphors of evil based on idea(l)s of moral and physical purity.

We are broadly describing this whispery phonatory profile and its use within this context as stances of 'threat' and argue that these stances can be employed regardless of whether the character represents human or non-human identities (i.e. this whispery phonation can be employed to express notions of non-human evilness, as well as 'threats' and evilness from human or human-like characters). The Allies typically employ stances of 'safety' (and by extension 'comfort', 'vulnerability', and 'trust') and this is realised through breathiness. In each case, these stances are either projected towards others or assumed upon themselves. For example, on the very rare instances where an Allied character employs whisperiness, they are doing so to signal

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

some threat is being put upon themselves or the party rather than themselves BEING a threat. While Enemy characters who employ whisperiness are signalling that THEY THEMSELVES are the existential threat.

As previously mentioned, our acoustic analyses collected measurements for F0, CPP, H1*-H2*, H1*-A1*, and HNR35. The data was first visually inspected in R (R Core Team 2023). Characters were coded for Alignment (Ally, Enemy, Neutral), gender (female, male, non-gender), Emotion (Neutral, non-Neutral), and characters ($n = 22$). The visual inspection revealed quasi-complete separation linked to gender. Namely, non-gender characters are always Enemies (i.e. no Ally or Neutrally Aligned non-gender character comparisons are available). Furthermore, non-gendered characters' speech is always non-Neutral Emotion, that is, no non-Neutral Emotion is available for non-gendered characters. However, the visual inspection did not suggest any patterns that could be explained through interactions of these variables.

Linear mixed-effects models were employed using the R software (R Core Team 2023) with the lme4 package (Bates, Mächler, Bolker, & Walker 2015). An estimation of 95% confidence intervals was computed through 5,000 parametric bootstrap replicates (Bates et al. 2015) implemented within the boot package (Davidson & Hinkley 1997; Canty & Ripley 2024). While bootstrapping models do not yield *p*-values, the resulting output offers an alternative measure of the reliability of the effect, replacing traditional significance values. This output can be interpreted such that if the confidence interval excludes zero, the effect is considered a robust predictor of the observed variation, or 'statistically significant'.

We note that although our models aim to encompass the full picture of our dataset, the nature of D&D introduces imbalances across three factors. Some characters only exhibit both Emotional Contexts after taking an Alignment shift into account, and within our dataset, character race/creature type is often categorically correlated with morality. Additionally, again, all non-gendered characters are limited to Enemy Alignment. Given this, it is not just Emotional Context, or race/creature type, or gender which is confounded with character morality, but rather, everything is. These confounds may reduce the statistical reliability of our modelling. Nevertheless, the trends outlined below align logically with previous literature on voice quality (see *Vocal profiles* below). With these considerations in mind, our models include three variables comprising four fixed effects: Gender, Alignment, and Emotion with an interaction effect between Alignment and Emotion. Character is set as a random intercept. The reference level for each model is Gender = Male, Alignment = Neutral, Emotion = Neutral. Separate models were produced for CPP, F0, H1*-H2*, H1*-A1*, and HNR. In the interest of space, the results of our acoustic analyses presented below are for F0 and CPP. Results for H1*-H2*, H1*-A1*, and HNR35 follow similar patterns as CPP. Figure 5 presents results for the model 'lmer(F0 ~ Gender + Alignment * Emotion + (1|Character)'. Similarly Figure 9 shows results for the same model for CPP.

Comparisons between the Neutrally Aligned non-Neutral Emotion and non-Neutral Emotion for Ally/Enemy characters are not presented within the model since the Neutrally Aligned non-Neutral Emotion comprises a single

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

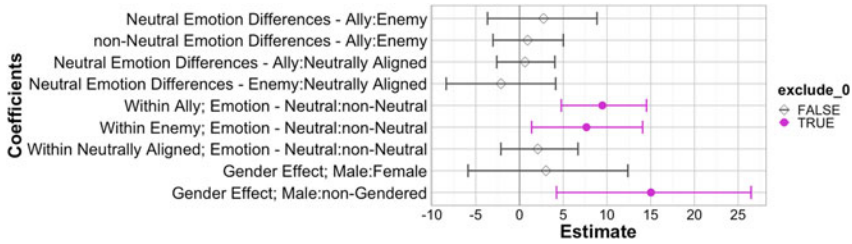

Figure 5. F0: 95% bootstrapping confidence intervals. Statistically reliable effect in magenta (CI's excluding 0 = TRUE).

character. Furthermore, while we find character race/creature type potentially relevant in a descriptive sense, there is too much (co)variation, and with unequal representation of individual levels, that the inclusion of this factor within the statistical models would not provide anything interpretable. We therefore do not include this within our statistical models, nor do we discuss race/creature type within our analysis beyond broad notions of human(esque) and non-human characters.

The model for F0 in Figure 5 indicates no differences between male and female characters, nor any differences between how Ally, Enemy, and Neutrally Aligned characters employ F0 variation based on Emotional Context. While Enemies overall may have a slightly lower F0 than Neutrally Aligned characters, this difference is negligible. We do, however, see results wherein non-gendered characters appear to employ a higher F0 than the other characters of the study. It is quite possible that this result emerges out of the combination of two factors: (i) Timorei (a Somnovem subpersona) is among the characters within the dataset with the highest F0 values, and (ii) Ira (another Somnovem subpersona) has a high probability of failure with regard to F0 tracking due to the nature of his voice quality being almost exclusively extreme growl (as discussed below). This would mean that this result comes about from the data being skewed higher than should be expected, and the descriptive statistics help illustrate this (Figure 6). Given these results we can confidently say that while Mercer may signal gender or Alignment at various points, mean F0 is not typically a resource employed to this end. Impressionistically, F0 dynamism may vary more so according to gender and Alignment with females and Allies being more variable overall; however, a quantitative analysis of pitch dynamism is beyond the scope of the present study. Interestingly, our results also indicate that WITHIN the Ally and Enemy characters, Mercer employs a higher mean F0 when in a non-Neutral Emotion than in the Neutral Emotion. This may be a result of the markedness of the non-Neutral Emotional Context. It would make sense for something emotionally marked to be linked with something vocally marked. This is corroborated with the results for Neutrally Aligned characters being generally unmarked across all measures, suggesting that Neutrality, in either Emotion or Alignment, is realised via unmarked speech acts.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

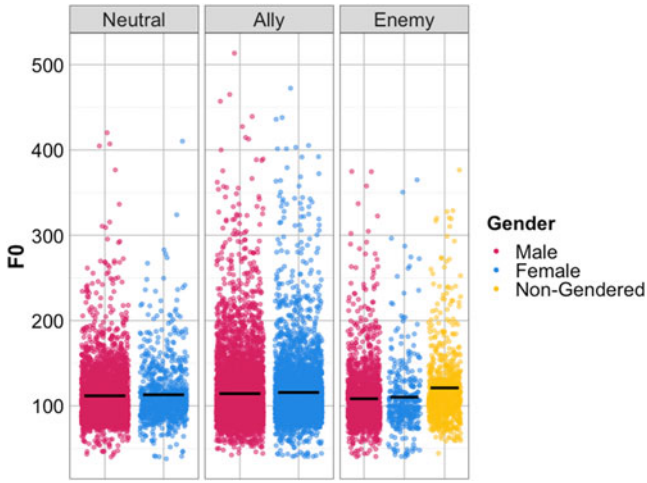

Figure 6. Gender results of F0 for all characters faceted by Alignment.

In terms of the CPP results, CPP should generally be realised within the acoustic measurements on a continuum from modal to breathy to whispery, based on whisperiness involving more turbulent and intense friction than breathiness (Esling et al. 2019:56, 124). Because lower CPP indicates a more breathy phonation when compared to modal voice, the lowest CPP is expected for whispery phonation.

Figure 7 shows the overall CPP results for all characters based on their Alignment, and the CPP continuum mentioned above is generally reflected in these results. Neutral characters are mostly modal (occasionally employing breathiness, and whisperiness is only employed in one extreme circumstance, explained below). Allies tend to exhibit modal phonation but will often draw on breathiness in more emotionally charged contexts.

The fact that breathiness is correlated with being an Ally and generally friendly affective stancetaking is not surprising. Breathiness has shown to be associated with intimacy and friendliness (e.g. Noble & Xu 2011; Starr 2015). Starr and colleagues (2020) show that ASMR (Autonomous Sensory Meridian Responses) performers in China typically employ breathiness to establish an intimate relationship with their audiences. In contrast, Enemies tend to be either fully voiced in whispery phonation, or they are relatively modal and employ (voiceless or voiced) whisper in more emotionally charged contexts. These patterns become evident when looking at CPP results through the lens of Emotional Context as in Figure 8. In all cases, Neutral Emotion is realised as modal phonation. However, it is within the non-Neutral Emotion that sociolinguistically relevant variation occurs. In short, this corresponds to breathiness for the Allies and whisperiness for Enemy characters.

The constantly Neutral characters are overall rather modal, displaying intermediate profiles for breathiness and pitch dynamism. The Neutral Aligned

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

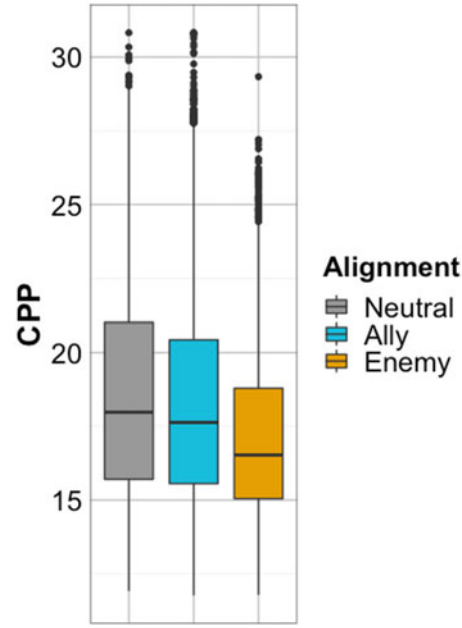

Figure 7. Overall CPP for all characters based on Alignment.

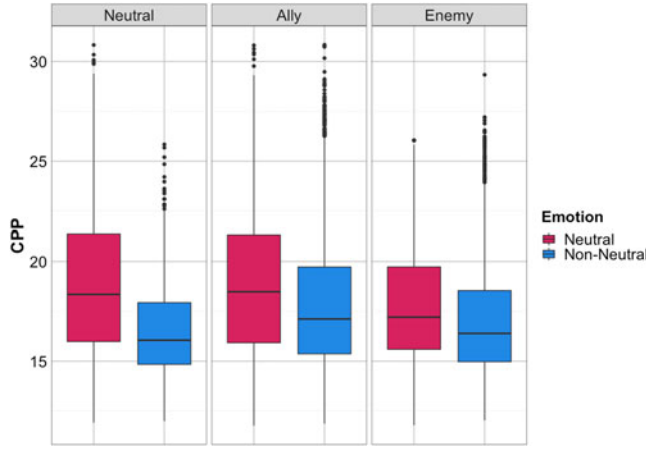

Figure 8. CPP for all characters by Emotional Context facetted by Alignment.

characters' non-Neutral Emotion is entirely derived from a conversation with the corpse of Vess DeRogna which is realised as extreme whisperiness. At no other point do the Neutral Aligned characters exhibit non-Neutral Emotion, nor at any other point do the Neutrally Aligned characters exhibit whisperiness.

The findings discussed here are supported by the results of our mixed-effects models for CPP (Figure 9). Allies show reliably lower CPP measurements

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

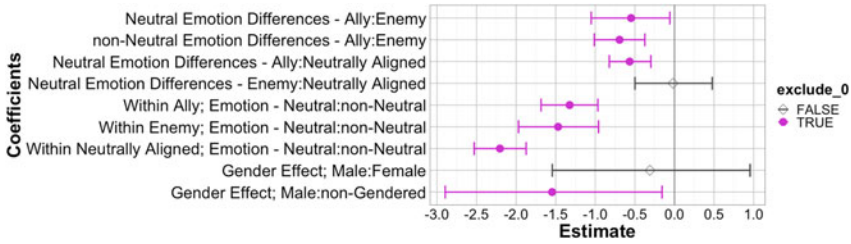

Figure 9. CPP: 95% bootstrapping confidence intervals. Statistically reliable effect in magenta (CI's excluding 0 = TRUE).

when compared to Neutral characters in both the Neutral and non-Neutral Emotional Contexts, which indicates some degree of breathiness even when in the Neutral Emotion. Allies further produce a lower CPP in the non-Neutral Emotion than they do in the Neutral Emotional Context. We also see a difference between the Allies and Enemies across both Emotional Contexts, wherein Enemies reliably produce lower CPP than the Allied characters. Furthermore, like Allies, Enemies produce a reliably lower CPP in the non-Neutral Emotional Context than in the Neutral Emotional Context. No differences are seen between Enemy characters and the Neutrally Aligned characters in the Neutral Emotional Context. Lastly, the gender effect reported within these models is likely a byproduct of all non-gender characters being evil/Enemy Aligned (which is largely characterised by whispery phonation).

Voice quality and performative stancetaking

As mentioned above, CPP should follow a continuum where modal voice has the highest of the CPP measurements, followed by breathiness, and then whisperiness. Unfortunately, the current machinery used for acoustic analyses of voice quality has proved not to be adequately equipped to handle extreme phonatory variation, at least not with respect to the vocal profiles analysed here (such as extreme whisper resulting in growling) and with the measures used here. Overall results presented in Figures 7 and 9 may indicate trends along the modal/breathy/whispery continuum corresponding to Alignment status; however, those results are enhanced by individuals who are saliently evil, always employing (extreme) whispery phonation while taking unambiguously threatening stances. These results are therefore complicated by several instances within our dataset where breathy and whispery profiles are shown to have nearly identical acoustic measurements across the measurements employed here.

Figure 10 illustrates the point that if our analysis were to rely solely on our acoustical measurements the differences between breathiness and whisperiness would be indistinguishable when looking at more emotionally complex characters. Here we show two characters, one Ally and one Enemy, each producing breathiness and whisperiness respectively. These characters are

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

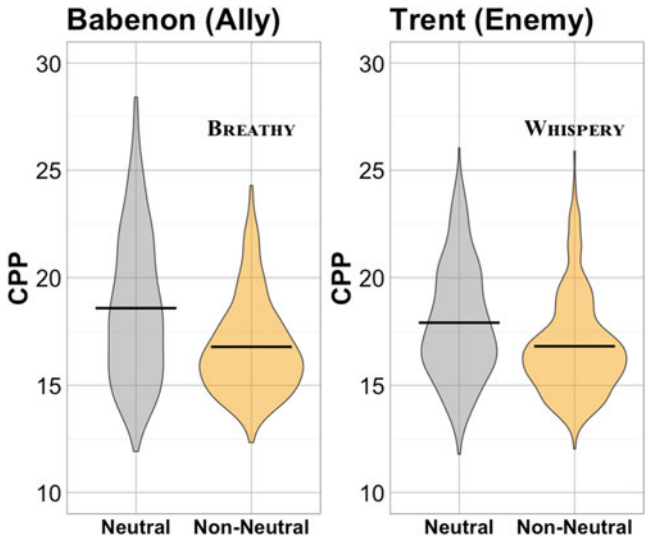

Figure 10. Babenon and Trent CPP by Emotional Context (line indicates mean value).

producing very different voice qualities, and are signalling very different things. Babenon is showing his love and affection for his daughter (a stance of safety), while Trent is performing acts of manipulation and showing a sociopathic malice through subtle threats of violence (a stance of threat). These two morally diametrically opposing speech acts are acoustically identical with regard to these measures but without an auditory analysis the social meaning would be lost. It is only when the acoustic and auditory data are examined in close conjunction can this difference be identified, allowing us to piece together the sociolinguistic meaning of the two phonatory profiles which index a characters' moral leanings.

Results of our voice quality analysis align with observed patterns of breathiness and whisperiness in character performances elsewhere. Heroes (or Allies in our context) commonly employ breathiness, whereas villains (or Enemies) rely on whisperiness (cf. Teshigawara 2003; Starr 2015). Sounds less common in everyday interactions, such as extreme whisperiness (which may even reach growl—similar to Callier 2013's 'harsh voice'), are almost exclusively employed by Enemies, especially when conveying stances of 'threat'. This taps into an aversion to 'otherness', particularly if that 'other' suggests something non-human-like (Kjeldgaard-Christiansen et al. 2023). As Kiesling (2022:412) suggests, 'how... and which stances are taken are often habitually repeated by people with similar identities', and we see this in Mercer's character personae across the good/evil spectrum.

While evilness is conveyed via whisperiness, Mercer draws on other vocal features to signal existential threat to the players as well. For example, he combines whisperiness with features associated with pathologically affected voices (e.g. Isharnai shows features associated with the ageing voice, and with a

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

pathologically ageing voice in particular; Baken 2005). Mercer also employs whisperiness in conjunction with a pulmonic ingressive airstream as a way to signal non-human-ness while voicing the Somnovem hive-mind. For characters like Isharnai and the Somnovem, these features are exclusively employed in conjunction with whispery phonation. In fact, non-human characters more broadly almost exclusively employ whispery phonation. Ultimately, the more whispery your whisperiness is, the more likely you are to be an Enemy, to be evil—to be a threat.

Our argument that stances of 'threat' can be exhibited via whispery phonation is something that Mercer is, at least on some level, aware of. During episode 46, the Allied character Orly Skiffback tells the party that they're sailing "into a proper storm". Mercer is then directly prompted by a player: "Damn. Could you be more ominous Orly?". At this point Mercer repeats himself but moves from what is initially a slight whisperiness to a noticeably increased whisperiness (Figure 11). This shift exhibits the fact that performed stances of 'threat' are realised with an increase in whisperiness and this increase in whisperiness explicitly signals to the players that they are entering potentially dangerous waters.

Somnovem, the hive-mind, presents us with a more nuanced approach to these patterns by allowing us to look at subsets of a single entity who portrays varying degrees and manifestations of evil-ness (Figure 12). Somnovem here represents instances where the whole hive-mind speaks as one, where the other three are individual minds of the hive that speak to the players separately from the mind. For the most part, these varying degrees of evil-ness are reflected in the acoustic measurements. Gaudius is the least evil entity of the hive-mind that the players and audience meet, offering a (warped) representation of 'love'. Gaudius employs a mostly modal phonation, with occasional lapses into breathiness. The breathiness is particularly enhanced when discussing 'love and family'. Timorei manifests as a symbolic representation of 'fear'.

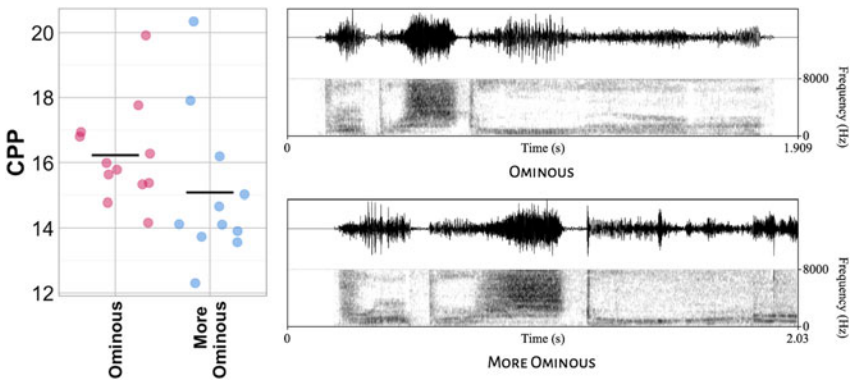

Figure 11. CPP results for Orly Skiffback stating "Capt'n, I think we're headin' into a proper storm" before and after being prompted to "be more ominous". Spectrogram shows Mercer stating the phrase "Proper storm".

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

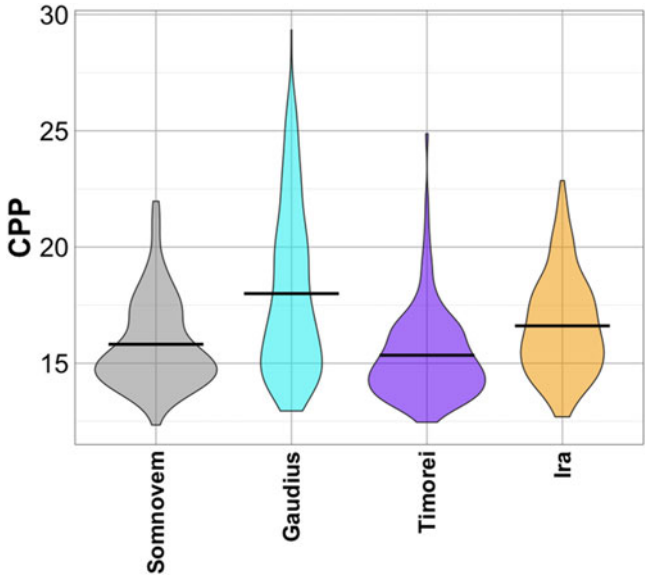

Figure 12. CPP by individual 'minds' of the 'Somnovem hive-mind' (line indicates mean value).

The speech pattern of Timorei is erratic, displaying a very harsh whisper. Mercer comments that in communicating with Timorei the players are witnessing 'a shotgun of fear and communication through a veil of scattered insanity and madness' (Critical Role LLC 2021). This patterns as we would expect with CPP, wherein Gaudius being modal with an occasional breathiness has a higher CPP than the other three who employ varying degrees of whisperiness throughout their entire speaking time. Timorei's consistent harsh whisper is more extreme than that of the greater Somnovem hive-mind, resulting in lower CPP. Where the acoustic measures fail to meet our expectations is in regard to Ira, a manifestation of 'wrath'. Auditorily, Ira has one of the harshest whispers of any of our characters. The extreme whisperiness is so prevalent and intense it is almost always realised as growl. However, that extreme and intense whisperiness is not represented as we might first expect in these acoustic measurements, showing CPP values higher than that of Somnovem. Our findings here may suggest that the quasi-periodic nature of growl may affect the CPP measurements. This finding would not be evident if we relied solely on the acoustic measures employed here.

Changing alignment, changing stance, changing voice

Four characters within our dataset change their Alignment through the campaign. These changes are reflected in their vocal performance as well. Lucien draws on a combination of breathier phonation and modality while Neutrally Aligned to the party, but he becomes more modal or whisperier when clearly

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

an Enemy, exhibiting a lack of breathiness previously displayed. Interestingly, Lucien's phonation is most unambiguously whispery not only when he explicitly threatens the players, but also when he explicitly discusses the Somnovem and his end goals (which will reign death and destruction upon the world). The main phonatory changes are therefore also employed to signal affiliation to parties other than the players themselves. In this way, Mercer employs phonatory variation to take both direct and indirect stances of threat, both of which signal Lucien's intent and moral character.

In addition, Lucien provides us with a performed, or fake, alliance. Where this happens, although the character is portrayed with the use of breathiness, this breathiness occurs abruptly, whereas in contexts where it is clearer that his alliance is genuine, the presence of breathiness is more stable and more gradiently varied within an utterance. This falls in line with Starr (2015:19), who shows that spoiled (villainous) princesses adopt features of feminine linguistic behaviour, but the lack of sweet voice (associated with certain sympathetic characters) reveals the pretence: 'While she can use grammatical features to put on a front of false femininity, she lacks a sweet voice quality… No matter how she disguises her personality, her unpleasant internal character prevents her from having a sweet voice'. This inauthenticity is indeed something we see in Lucien's fake alliance—he adopts breathiness at times, but this breathiness is not the same as the breathiness of the genuine allies. As Starr (2015:20) suggests that 'sweet voice cannot be faked; one must be an authentically sweet person to have a sweet voice', we too suggest that a character cannot accurately fake alliance or moral honour. This may be evidenced by the lack of trust the players put in Lucien as well. Despite Lucien's momentary Neutral alignment to the party, his instances of more abrupt breathiness do nothing to assuage the player's concerns that his INTENT is nefarious.

Yussa provides a contrast to Lucien, changing Alignment from Neutral to that of an Ally. Their phonation becomes breathier and shows more pitch dynamism after the shift. When Neutrally Aligned, they are nevertheless slightly breathy and do show some pitch dynamism, both of which seems to fall between those of Allies and Enemies.

Initially intended as an antagonist, Essek Thelyss begins as a Neutrally Aligned character and exhibits a considerable degree of breathiness (for a Neutral character), alternating with modal voice between and within utterances and accompanied by moderate pitch dynamism. Over time, as the Mighty Nein forms a friendship with Essek, Mercer's original plans for the character evolve, and Essek becomes an Ally. When the party uncovers Essek's past scheme that triggered an earlier conflict, they confront him about his actions. During this open discussion, Essek's phonation becomes noticeably breathier, signalling friendliness, while a decrease in pitch dynamism reflects the admission of guilt and the emotionally charged context. As a lasting friendly atmosphere is established between Essek and the party, his vocal profile becomes overall breathier and displays a relatively high pitch dynamism. Essek is also the only character who employs (lax) creak for any meaningful amount of time and he does so as a form of negative

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

stancetaking (which occurs primarily across single utterances irrespective of Alignment status).

The last character with a change in Alignment is Avantika, who turns into an Enemy from a former Ally. In many ways Avantika may be viewed as a classic femme fatale, manipulating the party to her whim, having a brief romance with one of the players in an attempt to gain their trust, but ultimately having her own interests at heart. During her time as an Ally, Avantika displays high degrees of breathiness and pitch dynamism, the former of which is in line with femme fatale vocal portrayals (Hejná & Eaton forthcoming). This breathiness is particularly accentuated during the aforementioned romantic encounter with the player character. Her phonatory patterns change when the party betrays her after discovering her true intentions. She shifts to more modal phonation with limited pitch dynamism. This shift from breathier to modal and even tense phonation is in line with constricted laryngeal settings being employed for negative affective stancetaking while simultaneously displaying her contempt and anger at the party for betraying her.

Even momentary changes in affiliation and affective stancetaking manifest themselves frequently in the vocal performance. One such example is the point when Isharnai, an Enemy 'hag' (a malevolent, supernatural creature associated with dark and sinister magical practices and who take on forms of, often grotesque, old women), consumes a cupcake sprinkled with magical dust which makes her forget about the highly conflictual relationship with the players. Having ingested the enchanted cupcake, Isharnai's vocal profile becomes breathier and her pitch dynamism also momentarily increases, drawing on resources that indicate she is less antagonistic and (at least momentarily) no longer a threat.

One character has a dramatic shift in voice quality which occurs not out of a change in Alignment, but a change in her overall state of existence. The only time whisperiness is seen within the Neutral characters is specifically while the party is speaking to the corpse of Vess DeRogna. This happens following her death when a player at the table casts the spell 'Speak with [the] Dead', which allows players to converse with a deceased through their former corpse. This too follows the expected phonatory pattern exhibiting a dramatic increase in whispery phonation (as seen in Figure 9). Corpses are typically associated with physical decay. As such, they can be considered cases of extreme sickness, representing something no longer human (or elf in this case), and definitely 'other'. Our results illustrate that whisperiness is associated with bad atypicality so it should be unsurprising, and even expected, that a dead creature would whisper.

Conclusions

We have asked how NPC identities in D&D can employ socially recognisable features associated with voice quality to construct personae in high performance that convey aspects of character morality. More specifically, we investigated what (if any) consistencies can be identified across character vocalisation based on whether the character is good (Ally) or evil (Enemy) in relation to the players. Through a holistic analysis of voice quality in nineteen characters,

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

breathiness emerged as signalling positive morality and stances of safety, comfort, and trust, where whisperiness signals negative morality and stances of threat. Qualitatively, pitch dynamism was also found to correlate with morality and stancetaking: the more limited the pitch dynamism, the more likely it is that the character portrayed is an Enemy and that they adopt stances of threat. This is in line with research that shows that higher pitch dynamism is often evaluated as more friendly (e.g. Daly & Warren 2001). Importantly, these patterns function both across categorical differences in alignment but are also employed to reflect more nuanced differences. Thus, more evil characters produce more whisper than less evil characters, and where generally morally positive characters momentarily perform stances of threat, they too adopt whisperiness. The two primary settings of the Laryngeal Articulator Model (Esling et al. 2019), namely constricted and unconstricted larynx, are used for fundamentally different social functions, where tenseness and whisper are manifestations of constricted laryngeal settings and breathiness of unconstricted settings. This is as could be expected wherein constricted larynx is associated with more tension than unconstricted larynx, and we propose, in line with Esling et al. (2019:125), that this physicality of the two settings makes them particularly prone to indexing positive and negative affective stancetaking.

Ultimately, the application of recognizable social indexes conveyed through variation in voice quality not only aids in character portrayal of various 'characterological figures' but also serves as a performative act, where the linguistic features become tools for signalling moral stances. This concept of morality as performed stancetaking extends our understanding beyond the mere representation of characters, emphasising how voice quality becomes a dynamic and intentional performance tool within digital storytelling. In essence, our study not only aids in the understanding of linguistic strategies in performative roleplaying but also underscores the broader relevance of sociolinguistic research in diverse linguistic landscapes. As technology continues to shape how we interact with narratives and characters, our findings encourage ongoing exploration into the intricate ways language is employed to convey meaning and emotion in fictional and real-world contexts.

Notes

* Carlsberg Foundation Young Researcher Fellowship (35891)

[1] 'Actual-play' refers to a genre of media which broadcasts live tabletop role-playing games.

[2] Our data does not include any audio from *The Mighty Nein Reunited,* a two-part special which premiered on November 17, 2022.

[3] See https://www.critrolestats.com/.

[4] Though Uk'otoa failed to meet the third criterion, their importance to the plot was so substantial that the authors felt their inclusion was essential.

[5] Mercer has publicly stated outside of *CR* that Yussa Errenis is a trans-male individual. We include him within the male cohort of characters because this information never entered gameplay, Yussa uses *he/him* pronouns, and most audience members would be unaware of this information unless they sought it out. Furthermore, no differences are seen based on gender for any measure.

[6] We use 'creature type' throughout the article for ease of general understanding with the acknowledgement that 'creature type' has a specialised meaning in D&D (i.e. fey, fiend, etc.; see also Wizards RPG Team 2022).

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

References

Agha, Asif (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1):38–59.

Agha, Asif (2011). Large and small scale forms of personhood. *Language & Communication* 31(3):171–80.

Anikin, Andrey; Katarzyna Pisanski; Mathilde Massenet; & David Reby (2021). Harsh is large: Nonlinear vocal phenomena lower voice pitch and exaggerate body size. *Proceedings of Royal Society B* 288:20210872.

Ayala, Francisco J. (2010). The difference of being human: Morality. *Proceedings of the National Academy of Sciences* 107:9015–22.

Baken, Ronald J. (2005). The aged voice: A new hypothesis. *Journal of Voice* 19(3):317–25.

Bates, Douglas; Martin Mächler; Ben Bolker; & Steve Walker (2015). Fitting linear mixed-effects models using lmer4. *Journal of Statistical Software* 67(1):1–48.

Bell, Allan, & Andy Gibson (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics* 15(5):555–72.

Boersma, Paul, & David Weenink (2023). *Praat: Doing phonetics by computer*. Online: https://www.fon.hum.uva.nl/praat/.

Callier, Patrick (2013). *Linguistic context and the social meaning of voice quality variation*. Washington, DC: Georgetown University PhD dissertation.

Canty, Angelo, & Brian Ripley (2024). boot: Bootstrap R (S-Plus) Functions. R Package version 1.3–31.

Coadou, Marion (2006). Voice quality and variation: A pilot study of the Liverpool accent. In *Speech prosody, Dresden, Germany, May* 2–5, 2006. Online: https://www.isca-archive.org/speechprosody_2006/coadou06_speechprosody.pdf.

Coupland, Nikolas (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.

Critical Role LLC (2021). Welcome to Cognouza | Critical Role | Campaign 2, Episode 137 [video]. 10 May. 1:08:06. Online: https://www.youtube.com/watch?v=l_loCAlJIVs&t=4086s.

Daly, Nicola, & Paul Warren (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics* 5(1):85–96. Online: https://doi.org/10.1111/1467-9481.00139.

Davidson, Anthony, & David Victor Hinkley (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Eckert, Penelope (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41:87–100.

ELAN (2022). ELAN, version 6.4 [computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Online: https://archive.mpi.nl/tla/elan.

Esling, John H.; Scott R. Moisik; Allison Benner; & Lise Crevier-Buchman (2019). *Voice quality: The laryngeal articulator model*. Cambridge: Cambridge University Press.

Esposito, Lewis (2016). 'I am a perpetual underdog': Lady Gaga's use of creaky voice in the construction of a sincere pop star persona. Swarthmore, PA: Swarthmore College BA thesis.

Fraile, Rubén, & Juan Ignacio Godino-Llorente (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control* 14:42–54.

Garellek, Marc (2020). Acoustic discriminabilty of the complex phonation system of !Xóõ. *Phonetica* 77:131–60.

Garellek, Marc (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94:101155.

Gordon, Matthew, & Peter Ladefoged (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29(4):383–406.

Hejná, Míša, & Mark Eaton (forthcoming). "It's chiefly your eyes I think, and that throb you get in your voice": The place of creaky voice in the soundscape of attractive female voices in twentieth and twenty-first century American cinematography. In Francesco Venturi (ed.), *Creak: Theories and practices of pulse phonation*. Stanford, CA: Jenny Stanford Publishing, to appear.

Hejná, Míša; Pavel Šturm; Lea Tylečková; & Tomáš Bořil (2020). Normophonic breathiness in Czech and Danish: Are females breathier than males? *Journal of Voice* 35(3):498.e1–22.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

Hill, Alexander K., & David A. Puts (2021). Vocal attractiveness. In Todd K. Shackelford & Viviana A. Weekes-Shackelford (eds.), *Encyclopedia of evolutionary psychological science*, 8441–45. Cham: Springer.
Jaffe, Alexandra (2009). *Stance: Sociolinguistic perspectives*. New York: Oxford University Press.
Kagan, Jerome (2000). Human morality is distinctive. *Journal of Consciousness Studies* 7(1–2):46–48.
Keating, Patricia; Christina Esposito; Marc Garellek; Sameer ud Dowla Khan; & Jianjing Kuang (2011). Phonation contrasts across languages. In *Proceedings of 17th International Congress of Phonetic Sciences, Hong Kong*. Online: https://www.reed.edu/linguistics/khan/assets/Keating%20ea%202011%20Phonation%20contrasts%20across%20languages.pdf.
Kiesling, Scott F. (2018) Masculine stances and the linguistics of affect: On masculine ease. *NORMA* 13(3–4):191–212.
Kiesling, Scott F. (2022). Stance and stancetaking. *Annual Review of Linguistics* 8:409–26.
Kelterer, Anneliese, & Barbara Schuppler (2019). Acoustic correlates of phonation type in Chichimec. In *Proceedings of Interspeech 2019, Graz,* 1981–85. North Hollywood, CA: Casual Productions.
Kjeldgaard-Christiansen, Jens; Míša Hejná; Mathias Clasen; & Mark Eaton (2023). Evil voices in popular fictions: The case of *The Exorcist*. *Journal of Popular Culture*. Online: https://doi.org/10.1111/jpcu.13234.
Laver, John (1980/2009). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Lawson, E. D., & Howard Giles (1973). British semantic differential responses to world powers. *European Journal of Social Psychology* 3:233–40 [Reprinted in *Peace Research Reviews* 6:25–38].
Lippi-Green, Rosina (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). Routledge.
Mackay, Daniel (2001). *The fantasy role-playing game: A new performing art*. Jefferson, NC: McFarland.
McAuliffe, Michael; Michaela Socolof; Elias Stengel-Eskin; Sarah Mihuc; Michael Wagner; & Morgan Sonderegger (2022). Montreal forced aligner [computer program]. Version 1.0.0. Online: http://montrealcorpustools.github.io/Montreal-Forced-Aligner/.
Mendoza-Denton, Norma (2011). The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology* 21(2):261–80.
Moisik, Scott R. (2012). Harsh voice quality and its association with blackness in popular American media. *Phonetica* 69:193–215.
Moisik, Scott R.; Míša Hejná; & John H. Esling (2019). Abducted vocal fold states and the epilarynx: A new taxonomy for distinguishing breathiness and whisperiness. In *Proceedings of 19th International Congress of Phonetic Sciences, Melbourne,* 220–24. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_269.pdf.
Murton, Olivia; Robert Hillman; & Daryush Mehta (2020). Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology* 29:1596–1607.
Noble, Lucy, & Yi Xu (2011). Friendly speech and happy speech: Are they the same? In *Proceedings of 17th International Congress of Phonetic Sciences, Hong Kong,* 1502–1505. Online: https://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Noble_Xu_ICPhS2011.pdf.
Ohala, John J. (1995). The frequency code underlies the sound symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols, & John J. Ohala (eds.), *Sound symbolism*, 325–47. Cambridge: Cambridge University Press.
Pisanski, Katarzyna, & David R. Feinberg (2018). Vocal attractiveness. In Sascha Frühholz & Pascal Belin (eds.), *The Oxford handbook of voice perception*, 1–24. Oxford: Oxford University Press.
Podesva, Robert (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4):478–504.
Podesva, Robert, & Patrick Callier (2015). Voice quality and identity. *Annual Review of Applied Linguistics* 35:173–94.
Pratt, Teresa (2023). Affect in sociolinguistic style. *Language in Society* 52(1):1–26.
Puts, David A.; Steven J. C. Gaulin; & Katherine Verdolini (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution & Human Behavior* 27:283–96.
R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Online: https://www.R-project.org/.

https://doi.org/10.1017/S0047404525000053 Published online by Cambridge University Press

San Segundo, Eugenia; Paul Foulkes; Peter French; Phillip Harrison; Vincent Hughes; & Colleen Kavanagh (2018). The use of vocal profile analysis for speaker characterization: Methodological proposals. *Journal of the International Phonetic Association* 49(3):353–80.

Shoaps, R. (2009). Moral irony and moral personhood in Sakapultek discourse and culture. In Jaffe, 92–118.

Shue, Yen-Liang; Patricia Keating; Chad Vicenik; & Kristine Yu (2011). VoiceSauce: A program for voice analysis. In *Proceedings of 17th International Congress of Phonetic Sciences, Hong Kong,* 1846–49. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Shue/Shue.pdf.

Starr, Rebecca Lurie (2015). Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society* 44(1):1–34.

Starr, Rebecca Lurie; Tianxiao Wang; & Christian Go (2020). Sexuality vs. sensuality: The multimodal construction of affective stance in Chinese ASMR performances. *Journal of Sociolinguistics* 24:492–513.

Stuart-Smith, Jane (1999). Glasgow: Accent and voice quality. In Paul Foulkes & Gerard Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 203–22. Arnold: London.

Teshigawara, Mihoko (2003). *Voices in Japanese animation: A phonetic study of vocal stereotypes of heroes and villains in Japanese culture*. Victoria: University of Victoria PhD dissertation.

Waskul, Dennis, & Matt Lust (2004). Role-playing and playing roles: The person, player, and persona in fantasy role-playing. *Symbolic Interaction* 27(3):333–56.

Wizards RPG Team (2022). *Mordenkainen presents: Monsters of the multiverse*. Wizards of the Coast LLC. Online: https://wpn.wizards.com/en/products/mordenkainen-presents-monsters-of-the-multiverse.

Yanushevskaya, Irena; Christer Gobl; & Ailbhe Ní Chasaide (2018). Cross-language differences in how voice quality and f0 contours map to affect. *The Journal of the Acoustical Society of America* 144(5):2730–50.

Cite this article: Boyd Z, Hejná M (2025). The 'Critical Role' of voice quality in Dungeons and Dragons: A case study of non-player characters voiced by Matthew Mercer. *Language in Society* 1–26. https://doi.org/10.1017/S0047404525000053