JOURNAL ARTICLE

Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts

Margaret E. L. Renwick[1] and D. Robert Ladd[2]

[1] Program in Linguistics, University of Georgia, Athens GA, USA

[2] Linguistics and English Language, University of Edinburgh, Edinburgh, UK

Corresponding author: Margaret E. L. Renwick (mrenwick@uga.edu)

It is known that the mid vowel contrasts of Standard Italian distinguish few minimal pairs, may be lexically variable, and show some degree of phonological conditioning in certain varieties. As such, they are relevant to recent suggestions that phonemic contrast may be partial, gradient, or otherwise more cognitively complex than traditionally assumed. Production data and vowel height judgments from 17 speakers confirm that most have clear *phonetic* distinctions between higher and lower mid vowels. However, the lexical distribution of these vowels is variable, and (in some speakers) phonologically conditioned to some extent; and though phonological awareness for all speakers is broadly accurate, we also observe cases where production and speaker judgment fail to match, in part because individual speakers' productions are variable. This suggests that the somewhat marginal status of the Italian mid vowel contrasts resides in the link between phonetic categories and individual lexical items, not in any indistinctness of the phonetic categories themselves.

1 Introduction

Reports of phenomena such as marginal contrasts, near mergers, and incomplete neutralization have appeared in the literature for many years, but phonological theory has yet to deal with them in any coherent way. Only recently has it been suggested (e.g., Hall, 2012) that the notion of *contrast* on which the phonemic principle rests may require substantial reexamination, at least to acknowledge the complex set of factors affecting the relationships among pairs of sounds. This paper aims to contribute to this reexamination by investigating the distinctions between higher and lower mid vowels in Italian—distinctions that are *phonemic* by conventional criteria, but that on several dimensions appear weaker than conventional assumptions would suggest.

1.1 Marginal contrast

Some concept of contrast is central to all theories of phonology. In classical phonemic theory, contrast was essentially the only criterion involved in deciding whether two phonetically distinct sound types represented two different phonemes. So long as a single lexical or grammatical difference in a language depended on a given phonetic distinction, that distinction was deemed to be phonemic and the two sounds were assumed to contrast in the language's phonology.

Even during the heyday of classical phonemic theory, though, phonologists were aware of problems with this assumption. Prague School theorists devoted considerable effort to

Art. 19, page 2 of 29

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts

the problem of neutralization (Trubetzkoy, 1936, 1939), formulating among other things the idea of the archiphoneme. Anglo-American phonemicists were especially troubled by cases where otherwise valid statements of allophonic complementary distribution broke down in specific contexts (e.g., German [ç]/[χ] (Moulton, 1947); American English *bomb/ balm* (Bloch, 1941); Scottish English *side/sighed* (Abercrombie, 1979); Fries and Pike's coexistent phonemic systems (1949)). There were also attempts to make sense of the fact that some phonemic contrasts seem to do more work than others; Hockett (1966) defined what he called 'functional load' based on the number of lexical confusions that would result if a phonemic contrast were lost. Nevertheless, the assumption that phonological contrast is categorical—that two sound types in a given language either do or do not represent distinct phonemes—was widely held, summarized in the slogan, "once a phoneme, always a phoneme."

Since the 1960s, of course, little attention has been explicitly devoted to defining the phoneme, but the problem of contrast and the idea of functional load have continued to attract a certain amount of theoretical attention. Goldsmith (1995) presents a taxonomy of problem cases and a discussion of their implications, which acknowledges that increasingly asymmetrical distributions between two sounds weaken the degree of contrast between them. Hall (2009, 2012) uses probabilistic calculations based on phoneme co-occurrence rates, similar in spirit to Hockett's notion of functional load, to determine the pairwise contrastiveness of sounds; sounds that are more predictably distributed across a corpus are less contrastive. Recent work suggests that functional load is relevant to various aspects of the way phonological contrasts evolve and are perceived. For example, Wedel et al. (2013) show that contrasts with higher functional load are less likely to be lost on a historical time scale, while Hall and Hume (2015) show that functional load is related to perceptual effects whereby sounds that are less contrastive are also perceived as more similar. More generally, it is clear from a useful review article by Hall (2013) that what she calls intermediate phonological relationships are by no means rare.

Various recent studies have discussed specific cases of marginal or quasi-phonemic contrasts, including some of those that worried phonemic theorists in the 1930s and 1940s. Some of these cases arise because the distribution of two different sounds is almost, but not completely, phonologically conditioned. This clearly applies, for example, to the Scottish English *side/sighed* distinction, where paradigm uniformity effects in morphologically complex monosyllables like *sighed,* and lexical variability in monomorphemic disyllables like *spider,* disrupt otherwise valid statements of classical complementary distribution (Scobbie & Stuart-Smith, 2008). It also applies to the distinction between the central vowels /ɨ/ and /ʌ/ in Romanian, which is investigated in detail by Renwick (2014). Renwick shows that the historical complementary distribution of these sounds was disrupted by the adoption of Slavic and Turkish loanwords containing /ɨ/, which has led to the development of minimal pairs; however, she also shows that the two sounds remain in near-complementary distribution and that this relationship appears to affect listeners' ability to distinguish them.

Nevertheless, not all instances of marginal or intermediate contrast involve near-complementary distribution. Some exhibit what Trubetzkoy (1939), discussing the higher and lower mid vowels in French, called *besondere Intimität* 'particular closeness.' As Trubetzkoy noted, the *phonetic* difference between French /i/ and /e/ is no greater than that between French /e/ and /ɛ/, but it is apparent to any French speaker that the latter two sounds are somehow related in a way that the former two are not. Some phonological conditioning affects the distribution of the French mid vowels, including complete neutralization of the higher-lower distinction in non-final syllables, but the situation cannot be described as a matter of complementary distribution with a few systematic exceptions.

This appears even truer of the corresponding distinctions between higher and lower mid vowels in Italian; as Ladd puts it (2006, p. 16), "There is a special relation of partial similarity between [Italian] higher and lower mid vowels. Somehow these vowels do not contrast with each other as completely as most other pairs of phonemes." Other such cases include: Distinctions between diphthongs and sequences of vowels in hiatus in Romanian and Spanish (Chitoran, 2002; Hualde & Prieto, 2002); distinctions between palatal sonorants and sequences of non-palatal sonorant and palatal glide in Catalan (Recasens, 1990; Recasens et al., 1991; Recasens et al., 1995); and perhaps also the distinction between tense /æh/ and lax /æ/ in Northeastern American English words like *bad* and *sad* (Labov, 1989). Plainly, phonetic similarity plays some role in these cases, and orthography may play a role in others. But the *phonological* nature of particular closeness, and its implications for the concept of contrast, remain unclear.

A possible approach to understanding problems with the notion of contrast has been suggested by Kiparsky (2014). Kiparsky decomposes the basis of phonemic status into two separate components, which he calls 'contrastiveness' and 'distinctiveness.' Contrastiveness is a structural notion, and relates to whether the distribution of the sounds in question is contextually predictable; distinctiveness is a perceptual notion, and refers to whether native speakers regard the sounds in question as phonetically different. The classical definition of a phonemic distinction requires phonemes to be both contrastive (contextually unpredictable) and distinctive (heard as different from other sounds), while classically defined allophones are neither contrastive (because one can predict from the phonological context which sound manifests a given phoneme) nor distinctive (because native speakers are typically supposed to be unaware of the difference between allophones of a given phoneme). However, Kiparsky notes that it is possible for contrastiveness and distinctiveness to vary independently, as shown in **Table 1**.

This gives rise to two combinations of properties that are unexpected in classical phonemic theory. In what might also be called 'allophonic awareness,' sounds may be distinctive—'quasi-phonemes' like German [ç]/[χ]—without being contrastive; this may arise from (or may give rise to) situations of near-complementary distribution. In the converse case of 'near contrasts,' speakers may consistently make a phonetic distinction without being consciously able to perceive it. Such cases have been attested both in experimental studies of so-called incomplete neutralization (Dinnsen & Charles-Luce, 1984; Ernestus & Baayen, 2006; Port & Crawford, 1989; Slowiaczek & Dinnsen, 1985; Warner et al., 2004) and in studies of sound change in progress (Hay et al., 2015; Labov, 1994; Labov et al., 1972).

Kiparsky's concerns are primarily with diachronic change, and his discussion is compatible with (though it does not require) a traditional categorical notion of contrast. That is, quasi-phonemes and near contrasts might be seen as merely transitional stages between one phonological status and another. However, it seems clear that not all of the problems with the classical phonemic notion of contrast are the result of change in progress. Ladd (2006) and Renwick (2014) both emphasize that the Italian and Romanian cases they discuss are diachronically stable, and Hall (2009) unambiguously treats synchronic phonological contrast as a matter of degree. One way or another, that is, phonological theory has no obvious place for special relationships (such as 'closeness' and

|	contrastive	non-contrastive
distinctive	**phoneme**	**'quasi-phoneme'**
non-distinctive	**'near contrast'**	**allophone**

Table 1: Independent variability of distinctiveness and contrastiveness (Kiparsky, 2014, p. 82).

'near-complementary distribution') between phonemes, and getting beyond a categorical notion of phonological contrast is therefore important not just for historical phonology, but for phonological theory generally. As Scobbie puts it in his discussion of what he calls the "phonology-phonetics overlap," there is a need for quantitative "*new* data" [emphasis in original], in order to "reinvigorate the descriptive basis of many phenomena, as well as provoking deeper theoretical understanding of the broader picture of linguistic sound systems" (Scobbie, 2005, p. 14).

1.2 Italian mid vowels

Standard Italian is conventionally described as having seven vowel phonemes /i e ɛ a ɔ o u/ in stressed syllables, with high-mid vowels /e o/ and low-mid /ɛ ɔ/, as shown in (1). As in a number of other Romance languages, the distinction between the higher and lower mid vowels is neutralized in unstressed syllables, with only the higher vowels /e/ and /o/ occurring in this position. However, the evidence for a special degree of closeness between the higher and lower vowels goes well beyond the fact that the distinction can be contextually neutralized. Even in stressed syllables, where they are supposed to contrast, the mid vowel pairs /e ɛ/ and /o ɔ/ are not normally orthographically distinguished, and linguistic descriptions concur in treating these distinctions as marginal (Vincent, 1988, p. 281f). There are few minimal pairs (examples in (2)), indicating a low functional load for the contrast, and in some words the mid vowel height is known to be variable. Prescriptive works for native speakers (Gabrielli, 1956) devote much discussion to the correct pronunciation of <e> and <o>, yet practical works for non-native speakers acknowledge that ignoring the distinctions creates few problems (Rebora, 1958). For all of these reasons the relation between the higher and lower mid vowels seems a likely source of insight into the nature of phonological closeness.

(1)   The Italian vowel space (Rogers & d'Arcangeli, 2004)
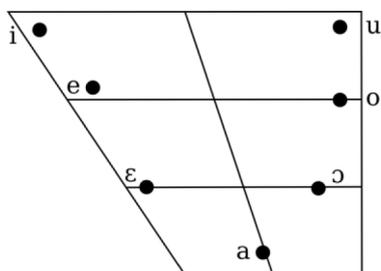
(2)   Minimal pairs

/e/ vs. /ɛ/:   /ˈpeska/ *pesca* 'fishing'   /ˈpɛska/ *pesca* 'peach'
/o/ vs. /ɔ/:   /ˈforo/ *foro* 'hole'   /ˈfɔro/ *foro* 'forum'

We make no apology for relying on a notion of 'Standard Italian,' though we acknowledge the difficulties involved in defining *any* standard language. In particular, we reject the notion that Standard Italian is a purely artificial construct, spoken natively by no one, and that it is therefore somehow illegitimate to ascribe to 'Italian' properties that are shared by speakers from different regions. Bertinetto and Loporcaro (2005) present a regionally nuanced description of contemporary Standard Italian, but reiterate the contention of De Mauro (1976) that since the 20th century Standard Italian has been "the language of everyday communication for all social classes." Clearly, local dialects survive to a greater or lesser extent in different parts of Italy (Canepari, 1979, 1980, 1999), and clearly the Standard Italian spoken in different regions, like any standard language,

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 5 of 29

is characterized by identifiable local accents (e.g., Savino, 2012; Wieling et al., 2014). This variation is reflected in our results, but in our view it does not invalidate the study of contrasts in Standard Italian considered as a whole, nor would it justify restricting our sample to speakers from a single region or city. At most, regional variation represents one aspect of the overall status of the mid-vowel contrasts in what we confidently describe as the Standard Italian speech community. Our speakers are all educated users of Standard Italian and many of them do not speak (though they may understand) a local dialect.

Our aim, then, is not to investigate questions of dialectology or sociolinguistics, but to see whether there are generalizations to be drawn about the manifestations of phonological closeness in the behavior of individual speakers, regardless of where they come from. The heart of our methodology is to link individual speakers' pronunciations to their own judgments of mid vowel height, a technique also used for Catalan mid vowels by Nadeu and Renwick (2016). Our specific research questions included the following:

- Do speakers have seven acoustically distinct vowel qualities, or do the higher and lower mid vowels overlap in acoustic space?
- Do speakers share the same judgments of mid vowel height across words, or is there variation in which lexical items have high- vs. low-mid vowels?
- Is there evidence of phonological conditioning of the choice of higher or lower vowel?
- Are speakers aware of their own production? Can they reliably judge whether in a given word they use the higher or lower mid vowel?
- Are there clear cases of *miscategorization*, in which the speaker's judgment and production do not match? How can these be identified?

By structuring our study around these questions, which focus both on individual speakers' phonological awareness and on the phonetic details of their productions, we implicitly test the validity of Kiparsky's distinction between distinctiveness and contrastiveness. As we shall see, our results suggest that it is indeed possible for sounds in a language to be clearly distinct *as phonetic categories* and yet fail to be distinguished reliably and consistently in the way they are deployed in the lexicon and in signaling phonological contrasts. We suggest that this apparent dissociation between phonetic distinctiveness and lexical contrastiveness is an important component of phonological closeness.

Our methods are described in Section 2 of the paper, while the results follow in two parts: Section 3 captures the relationship between speakers' intuitions of vowel quality and its phonetic realization across lexical items, while Section 4 explores cases when vowel intuitions and acoustics do *not* match for speakers of Standard Italian. Section 5 presents discussion and conclusions.

2 Data collection and methods
2.1 Materials

Experimental materials consisted of a set of Italian lexical items, containing a stressed target vowel in one of five contexts: ˈC**V**.CV.CV (antepenultimate stress on open syllable); CV.ˈC**V**.CV (penultimate stress on open syllable in trisyllabic word); ˈC**V**.CV (penultimate stress on open syllable); ˈC**V**N.CV (penultimate stress on syllable with nasal coda); ˈC**V**C.CV (penultimate stress on syllable with non-nasal coda). We also included one pretonic context (the initial syllable of C**V**.’CV.CV words), where the mid vowel contrast is said to be neutralized. **Table 2** summarizes the six contexts and gives an example of each, together with the informal shorthand we use to refer to them.

	Context	Shorthand	Example word		
1	ˈC**V**.CV	DA-da	/ˈsɔdo/	*sodo*	'compact'
2	ˈC**V**N.CV	DAN-da	/ˈnɔnno/	*nonno*	'grandfather'
3	ˈC**V**C.CV	DAD-da	/ˈsordo/	*sordo*	'deaf'
4	ˈC**V**.CV.CV	DA-da-da	/ˈdɛtʃimo/	*decimo*	'tenth'
5	CV.ˈC**V**.CV	da-DA-da	/baˈlena/	*balena*	'whale'
6	C**V**.ˈCV.CV	(pretonic)	/teˈnute/	*tenute*	'holdings, property'

Table 2: Target vowel contexts.

To compare the mid vowels to the speaker's vowel space as a whole, we recorded words containing /i u a/ in addition to the mid vowels. For each of the five stressed contexts, we included five items exemplifying all seven vowels; in the pretonic context there were five examples of all five possible vowels. This yielded a total of 200 target vowels. However, nearly all the words of the form da-DA-da were analyzed both for their stressed vowel and their pretonic vowel, so that the word list contained only 180 unique words. (For example, the test word *balena* 'whale' provided data both for stressed /e/ and for pretonic /a/; see Appendix 1.) Each item was randomly embedded in one of five prosodically similar frame sentences (see Appendix 1), and the resulting list was presented to participants in random order.

To avoid exaggerating the variability of the mid vowels in our data, we used only test words with a single prescriptive norm, based on the pronunciation recorded in De Mauro (2000). For example, all the items we used to capture prescriptive stressed /ɛ/ are listed by De Mauro as having only a single pronunciation, with [ɛ]. All of the rather numerous items recognized by De Mauro as variable (e.g., *lettera, freno, schermo, sonno, folla, posto*) were excluded.

The set of test words samples evenly not only across prescriptive vowel categories, but also across a range of phonological contexts. The trisyllabic test words, with either penultimate or antepenultimate stress, allow us to test for vowel quality differences across stress positions and to evaluate impressionistic claims of pretonic neutralization to [e] and [o]. Among the disyllabic words, the variety of syllable structures was intended to investigate the possibility of phonological conditioning of the speaker's choice of mid vowel. In particular, we expected allophonic variation in the closed syllables (the DAN-da and DAD-da sets). Such conditioning is found in French, with higher mid vowels tending to occur in open syllables and lower mid vowels tending to occur in closed syllables (this is the so-called *loi de position;* Fagyal et al., 2006; Scherrer et al., 2015), and dialectological descriptions suggest that there are similar tendencies in Italian. In Romanian and Portuguese, meanwhile, postvocalic nasals have historically triggered raising, and there is clear evidence of nasal influence on vowel quality in the history of Italian as well (Hajek & Maeda, 2000).

2.2 Speakers

Participants were recruited based on availability. We recorded 17 speakers (3 male). All were educated native speakers of Italian, and had no reported hearing or speaking problems; most were between 20 and 40 years of age, with one in her mid-fifties. All additionally spoke at least one other language, and recording occurred in several locations, all outside of Italy: In the Phonetics Laboratory at the University of Oxford; in the recording studio at the University of Edinburgh; in the Phonetics Laboratory at Cornell University; and in the Center for Language Science at Pennsylvania State University. Before participating in the study, speakers gave their informed consent, via a procedure approved by

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 7 of 29

the Institutional Review Board at the University of Georgia. Participants completed a questionnaire asking their place of birth, where they grew up or resided, their parents' origins, their knowledge of other languages, and how often they used Italian. The amount of time each had spent outside of Italy varied widely, from several weeks (speakers ItF11, ItF12, ItF13, ItF14, ItF15, ItM2, ItM3 were visiting the United States for three months) to multiple decades (speakers ItF5, ItF6, ItF8 are long-term ex-pats but all use Italian in their daily lives). Except for the three long-term ex-pats and ItF2, all speakers were graduate or undergraduate students at the time of recording. Speakers were compensated for their time.

Table 3 lists speakers' gender, hometown, and dialect area. Dialect area is very broadly defined for most speakers as either North or Central, primarily with reference to the La Spezia-Rimini Line, the traditional boundary between northern and central Italian dialects (Rohlfs, 1972). 'Central' in our sample of speakers refers to Tuscany and Rome; the South and Sardinia are represented by only one speaker each, which eliminates the need for us to be more precise about the boundary between Central and South. We emphasize, however, that our study is not intended as high-tech dialectology, but is aimed at investigating the apparent phonological closeness between the mid vowels and understanding the manifestations of this closeness in individual speakers' productions and in their phonological intuitions.

2.3 Acoustic data collection and analysis

Three repetitions of stimuli were recorded and analyzed. Depending on the recording location, speakers either read stimuli from a paper list, or from a computer screen via PowerPoint. In the case of obvious speech errors unrelated to mid vowel height, speakers were asked to repeat specific sentences. All speakers were recorded in a sound-attenuated booth; digital recorders were used in all locations, with sampling at 44.1 kHz.

Speaker ID	Gender	Hometown (Regione)	Dialect area
ItF1	F	Roma (Lazio)	Central
ItF2	F	Udine (Friuli)	North
ItF3	F	Ravenna (Emilia Romagna)	North
ItF5	F	Vicenza (Veneto)	North
ItF6	F	(Veneto – Marche – Lazio)*	North/Central
ItF7	F	Roma (Lazio)	Central
ItF8	F	Cagliari (Sardegna)	Sardinia
ItF9	F	Montebelluno (Veneto)	North
ItF10	F	Savona (Liguria)	North
ItF11	F	Matera (Basilicata)	South
ItF12	F	Ardenno (Lombardia)	North
ItF13	F	Parma (Emilia Romagna)	North
ItF14	F	Bologna (Emilia Romagna)	North
ItF15	F	Arezzo (Toscana)	Central
ItM1	M	Arzignano (Veneto)	North
ItM2	M	Locate Varesino (Lombardia)	North
ItM3	M	Arezzo (Toscana)	Central

Table 3: Study participants: Gender, hometown, and dialect area.
*ItF6 moved several times during her childhood and adolescence.

Data were automatically aligned using SPPAS (Bigi & Hirst, 2012), whose dictionary was augmented to include all words from our list. Segmentation in the resulting Praat TextGrids (Boersma & Weenink, 2015) was checked by hand against waveforms and spectrograms, to ensure that the boundaries of each target vowel were correctly placed, based on either the onset and offset of F2, or at the point of major spectral change separating the target vowel from preceding and following consonants. Formant values (F1, F2) were automatically extracted at each target vowel midpoint, using the Burg algorithm as implemented in Praat. Formant tracking settings were visually checked in Praat before extracting each speaker's data, to reduce tracking errors. For male speakers, five formants were extracted with a ceiling of 4500 Hz; for women, typically four formants were extracted with a ceiling of 5000 Hz, although one female speaker's data were extracted with four formants and a ceiling of 4500 Hz. The output was hand-checked, and erroneous formant values were corrected. A small number of tokens were eliminated (e.g., due to speaker error or noise in the signal), resulting in a total of 10,161 tokens across all seven vowels (5,571 mid vowel tokens).

Vowel formant data were subjected to a Lobanov z-score transformation (Lobanov, 1971), which results in speaker-intrinsic, vowel-extrinsic, and formant-intrinsic normalization shown to preserve vowel-specific information, as well as regional or sociophonetic characteristics, while minimizing variation due to anatomical variation from body size or gender, for example (Adank et al., 2004; Chen, 2008; Nadeu, 2014).[1] Normalized data are used in all statistical analyses that pool speaker data.

In Section 4, the normalized acoustic data are categorized on the basis of a k-means clustering analysis combined with a quantitative characterization of outliers. In the interests of coherent exposition, we postpone detailed discussion of the statistical methods used in that analysis.

2.4 Judgments

Some time after the recording session (usually within two weeks), participants completed a questionnaire asking whether they used the higher or lower vowel in the 100 test words with stressed mid vowels. The questionnaire was sent to participants via email and returned the same way. The instructions (in Italian) were to, "indicate with an '*A*' if the stressed syllable in a word is open [*aperto* = low-mid], or with a '*C*' if it is closed [*chiuso* = high-mid]. Indicate uncertainty with ?." These descriptive adjectives for vowel quality are generally known by educated Italian speakers and are used for talking about the mid vowel contrast.[2]

---

[1] In this procedure, mean values and standard deviations of F1 and F2 are calculated for each formant (F), for speaker (i), using data from all vowels. The difference is calculated between the F1 or F2 value of each token (n), and the appropriate formant-specific mean; and the resulting value is divided by the formant-specific standard deviation. Once applied to all tokens, the output is centered around a mean of (0, 0), where for example a high front /i/ has a negative value for z-scored F1 and a positive value for F2, while a low central /a/ has an F2 near zero and a positive F1. The formula for this transformation is given in (3).

$$(3) \qquad z_{ni} = \frac{\left(F_{ni} - \mu\left(F_i\right)\right)}{\sigma\left(F_i\right)}$$

[2] An Associate Editor asks whether investigations such as ours can only be done with speakers who are literate in an alphabetic writing system. We recognize the importance of such concerns, but we think that the use of Italian speakers' metalinguistic awareness of mid vowel height is best seen as a methodological shortcut to investigating phonological intuitions. That is, we believe that there are many behavioral manifestations of phonological intuition (including rhymes, word games, and setting of tunes to texts) that could potentially be used to probe hypothesized structural relations like phonological closeness, and that these do not necessarily require alphabetic literacy. In a language community where rhyme is a known concept, for example, it is possible to ask even illiterate speakers if two words rhyme with one another. Unquestionably the existence of alphabetic literacy and the widespread awareness of mid vowel quality make the present investigation easier, but we do not believe that similar investigations are in principle impossible under other conditions.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 9 of 29

3 Results I: Judgments and the vowel space

In this section we describe the correspondence between speakers' judgments and productions to answer several of our research questions. Specifically, we consider whether judgments of vowel height are shared by speakers; whether speakers have seven distinct vowel qualities; and whether phonological conditioning affects vowel quality. We show that speakers are generally reliable judges of their own vowel use, while the analysis that follows in Section 4 identifies and discusses mismatches between speakers' judgment and production.

3.1 Variation in speaker judgments

We begin with a brief overview of interspeaker variation in the judgment of vowel height across lexical items. We quantify the amount of interspeaker agreement by subjecting the matrix of judgments (*aperto* vs. *chiuso*, or ?, for each mid vowel, for each speaker) to a Fleiss' Kappa (κ) test (Fleiss, 1971), as implemented in the *irr* package (Gamer et al., 2012) for R (R Core Team, 2000). When all 17 speakers' judgments are simultaneously compared, the resulting κ coefficient is 0.399 ($z = 47.4$, $p < 0.001$), which according to Landis and Koch's (1977) guidelines indicates moderate agreement (on a scale from 0 = 'poor agreement' to 1 = 'near perfect agreement'); looked at the other way, there is considerable variability, or disagreement, in mid vowel height judgments across speakers. For instance, there are only 14 words out of 100 for which all speakers agreed with each other, and with the prescriptive norm. There were no words for which speakers unanimously disagreed with the dictionary, but for two words (*conga* 'conga' and *sede* 'seat') only one speaker chose the prescriptive norm. Full details of individual speakers' judgments appear in Appendix 2.

To be sure, the variability in speaker judgments is by no means unstructured; in particular, as would be expected from descriptions of variation in Standard Italian (Bertinetto & Loporcaro, 2005), we find certain regionally driven patterns in judgments. For example, Central speakers agree with the prescriptive norm to a greater extent than Northern speakers. Moreover, merely removing the speakers from Sardinia and the South (ItF8 and ItF11) increases the κ coefficient from 0.399 to 0.507 (moderate agreement; $z = 52.5$, $p < 0.001$); this increase is driven by the fact that the 15 Central and Northern speakers share judgments for 33 of the 100 lexical items rather than only 14. Nevertheless, while our participants shared many intuitions of mid vowel height, particularly with other speakers from the same geographic area, the fact that agreement was only moderate shows that there is also considerable individual variation in judgments across lexical items. We also note here that variation does not imply uncertainty: The ? label was used only nine times in 1,700 judgments.[3]

3.2 Correspondence between vowel labels and phonetic realization

In order to plot data points in F1, F2 vowel space or to discuss variability in the realization of vowels, it is necessary to code each point as representing one vowel phoneme or another (e.g., either /e/ or /ɛ/). There are two obvious criteria for doing this: The prescriptive identity of the vowel, and the speaker's judgment about which vowel is used in a given word. For our purposes the second criterion is more informative. If we plot vowels according to their prescriptive identity and find that a given speaker's vowels are scattered across both the high-mid and low-mid clusters, it tells us only that the speaker's

---

[3] With regard to patterns in speaker judgments, we note that the overwhelming majority of words (29/33) for which all Central and Northern speakers share intuitions have prescriptive high-mid vowels. The use of ?, on the other hand, was not evenly distributed across speakers: 7 of the 9 '?' judgments came from speakers ItF8 and ItF3.

productions do not correspond to prescriptive norms, but reveals nothing about whether they can accurately judge their own productions. If instead we plot each data point on the basis of the speaker's judgment of whether it is high-mid or low-mid, then the resulting plots allow us to assess whether speakers are accurate in their own classifications. Distinct clusters of vowels judged high-mid and vowels judged low-mid will indicate that the speaker's judgment corresponds to phonetic reality; the basis of any divergence from prescriptive norms is then a separate question. Naturally, for speakers whose usage coincides with prescriptive norms *and* who are aware of their own usage—which is of course what a classical phonemic conception of contrast would predict—the two types of plots should be identical.

The differences between the two methods of labeling the vowel tokens are illustrated in **Figure 1** for two representative speakers, ItF15 and ItF10. On the left-hand panels, mid vowel tokens are labeled according to prescriptive norms (PrescriptiveQ), and on the right, the vowel tokens are labeled according to the speaker's judgment (SpeakerQ); 95% confidence ellipses surround data points that have the same label. For ItF15 we find a close correspondence among prescriptive norms, the speaker's judgment, and the speaker's actual productions, so there is little difference between the ellipses in the two plots. For ItF10, however, there are considerable discrepancies between her productions and prescriptive norms, especially for the front vowels /e/ and /ɛ/, and here the two plotting methods yield different pictures. ItF10 uses a fairly high-mid vowel in many words that have /ɛ/ according to the prescriptive norm, so that in the left-hand panel the ellipse for /ɛ/ is quite large and almost totally overlaps that for /e/. In the right-hand panel, on the other hand, nearly all tokens labeled [e] or [ɛ] by the speaker's own judgment are gathered in one cluster each. We can conclude that even though this speaker's front mid vowel productions do not match prescriptive norms, she is nevertheless a good judge of her own vowel quality. Consequently, plotting by SpeakerQ gives a more useful representation of how she uses her vowel space.
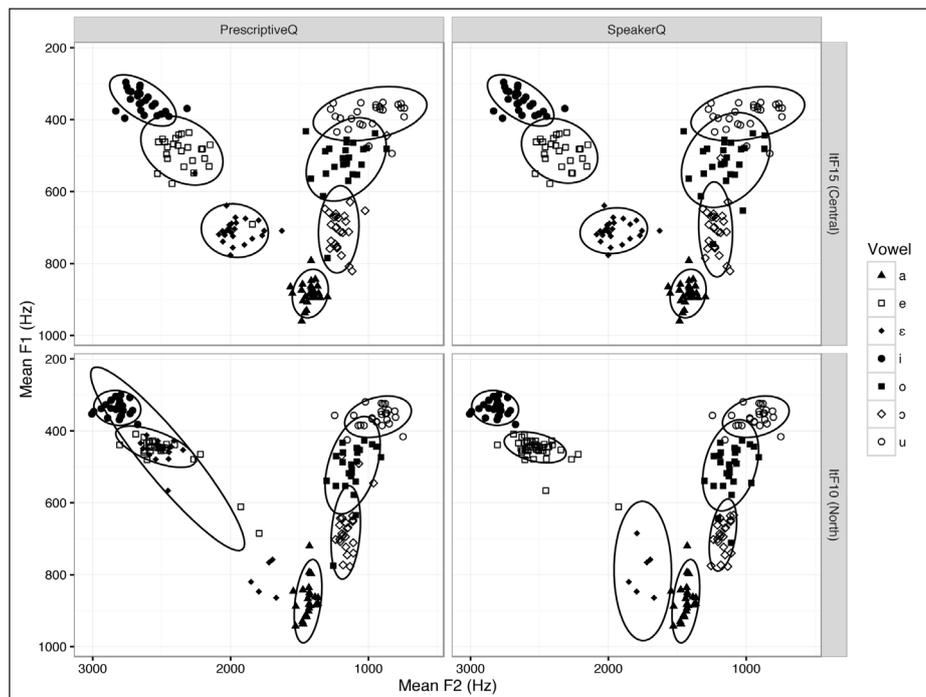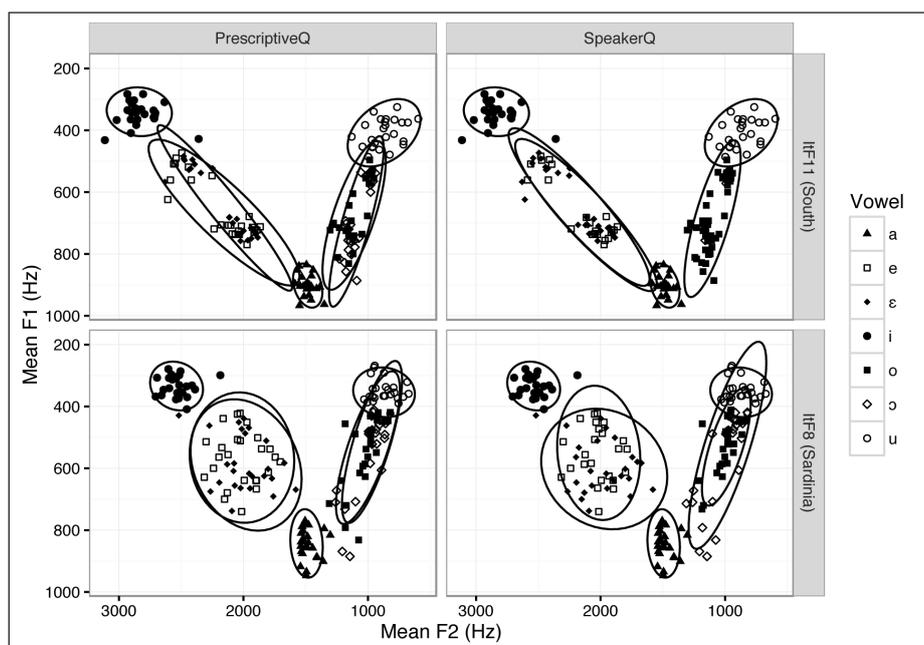

Figure 1: Stressed vowel space for speakers ItF15 and ItF10. Each point represents word averages of F1, F2. Left: Vowel quality (plotting symbol) assigned according to prescriptive norms (PrescriptiveQ). Right: Vowel quality assigned according to speaker judgments (SpeakerQ). Ellipses represent 95% confidence intervals.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 11 of 29

The impression that SpeakerQ is the basis of a more insightful analysis can be placed on a firmer quantitative footing. If speakers' self-reports match their pronunciation, but not the prescriptive norm, then we expect the variability of formant values in each vowel category to be smaller when vowels are classified according to SpeakerQ than when classified by PrescriptiveQ. Because the main difference between the high-mid and low-mid vowels is in vowel height, we focus on variability in F1 only. To evaluate it, the standard deviation of F1 values in the four mid vowels was compared across the SpeakerQ vs. PrescriptiveQ conditions using a linear mixed effects model, run in R using the lme4 and lmerTest packages (Bates & Maechler, 2009; Kuznetsova et al., 2013). The model, detailed in Appendix 3, shows a significant effect for each vowel type, and label is also significant ($p < 0.001$), indicating different F1 distributions of vowels when labeled according to SpeakerQ (lower standard deviation) vs. PrescriptiveQ (higher standard deviation). Thus supported by both data visualization and statistical modeling, throughout the Results section, mid vowel quality is plotted according to the individual speaker's judgment for each word.

Even when speaker judgments are used to plot data, however, there are two speakers whose results are unlike either of the patterns illustrated in **Figure 1**: ItF8 and ItF11. These two are exceptions to the generalization that speakers are reasonably accurate judges of their own usage; their data are plotted in **Figure 2**. Both apparently exhibit a complete absence of phonological awareness of mid vowel height, which means that the ellipses produced by the two plotting methods yield comparable pictures. However, the two speakers' acoustic patterns are quite different: ItF11 has four clearly separate clusters in acoustic vowel space, while ItF8 has only two rather diffuse mid clusters, one front and one back. ItF11 in fact judged so few words to have [ɔ] that no ellipse was drawn for that vowel in **Figure 2**.

These patterns are in fact expected on the basis of standard dialectological descriptions. ItF8 is from Sardinia and ItF11 is from Matera, in the far Southern part of the Italian

Figure 2: Stressed vowel space for speakers ItF11 and ItF8. Each point represents word averages of F1, F2. Left: Vowel quality (plotting symbol) assigned according to prescriptive norms (PrescriptiveQ). Right: Vowel quality assigned according to speaker judgments (SpeakerQ). Ellipses represent 95% confidence intervals.

peninsula. Sardinian speakers are generally described as having no distinction between higher and lower mid vowels (Jones, 1988), while in Southern varieties the higher and lower mid vowels are described as being in a strictly allophonic relation, conditioned by syllable structure and in some cases by other factors as well (Avolio, 1995). It is worth noting that the difference between these speakers draws attention to two possible outcomes of a diachronic merger: Merger can mean complete loss of phonetic distinction (which is what we apparently see in ItF8's data)[4], or it can involve maintenance of a phonetic distinction combined with completely consistent allophonic conditioning and lack of phonological awareness (which seems to be the case with ItF11).

Finally, although we have demonstrated that most speakers show good metalinguistic awareness of their own vowel productions, speaker judgments are by no means perfectly accurate. The plots in the SpeakerQ panels of **Figure 1** do exhibit mismatches, in which a token labeled as high-mid is clearly realized as a low-mid (or vice versa). Such mismatches, which are found to varying degrees for all speakers, are examined in more detail in Section 4.

3.3 Phonetic distinctness of vowel categories

In this section we quantify the amount of acoustic overlap among the vowel clusters. One conceivable manifestation of phonological closeness between mid vowels is that they occupy the same regions of acoustic space; in that case, we would expect to find less acoustic distance between mid vowels than between other pairs of vowels that are adjacent in the vowel space.
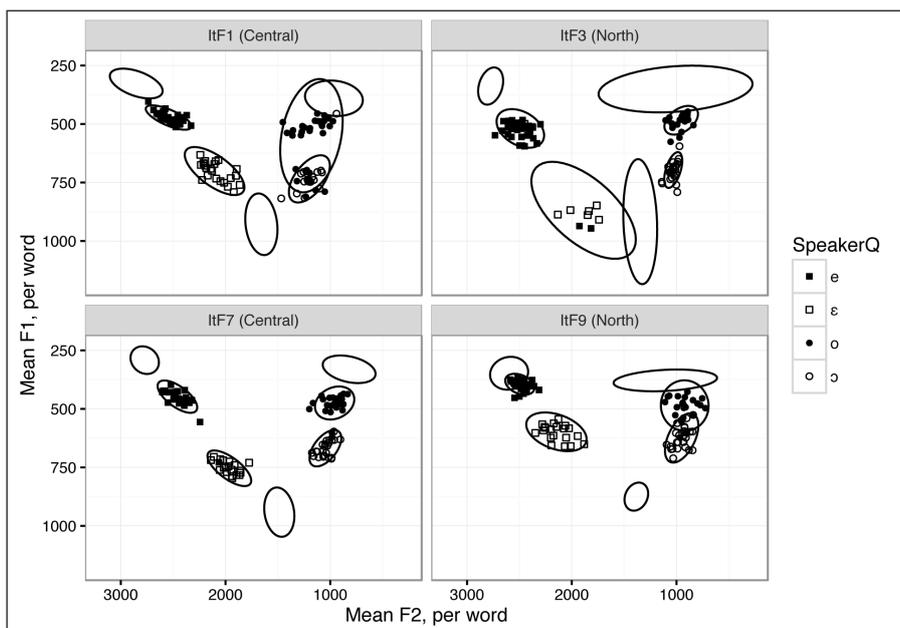
Visual inspection of vowel plots shows that with the exception of ItF8 (from Sardinia), our speakers all have seven distinct clusters of acoustic vowel quality (see individual plots in Appendix 4). **Figure 3** illustrates that speakers may have different degrees of phonetic separation between vowel categories: [ɛ, ɔ] occur at the same height for ItF1 and ItF9, while for ItF3 and ItF7 [ɛ] lies much lower in the vowel space. At the same time, the high-mid vowels for all speakers appear to be closer in F1 space to [i, u] than they are to the low-mid vowels. Most of the speakers show similar patterns, having overlap in the back vowels, but to differing extents, and with some overlap among all vowel categories.

Acoustic distance between mid vowels is summarized in **Table 4**, using both raw and z-scored values. This table shows, for each speaker, Euclidean distances between [e] and [ɛ] and between [o] and [ɔ]. With the exception of speakers ItF8 and ItF11, there are substantial distances between the higher and lower mid vowel centroids. This provides further support for the presence of seven distinct phonetic vowel categories in Italian, and makes it unlikely that acoustic overlap is the basis of the mid vowels' phonological closeness. Additionally, **Table 4** compares the F1 distances between front vowels to those between back vowels; these values are significantly different from zero (Hz: $t(16) = 5.7605, p < 0.001$; z-scored: $t(16) = 3.9121, p < 0.005$), indicating that front vowels are more distinct in F1 than are back vowels.[5]

However, in order to strengthen the conclusion that speakers have seven distinct vowel categories, it is appropriate to compare the distances between mid vowels to distances elsewhere in the vowel system. To do this, we calculated acoustic distance in the vowel

---

[4] Actually, it is likely that some Sardinian speakers show effects of metaphony, whereby mid vowels are higher if followed by a high vowel (/i/ or /u/) and lower elsewhere, as happens in most varieties of Sardinian (Bolognesi, 1998, pp. 19–22; Pittau, 1991, pp. 34–35). We were unable to test this systematically, because most of our test words, being citation forms, ended in /e, o, a/.

[5] This result includes all speakers, even ItF8 and ItF11. If they are excluded due to their lack of high-mid vs. low-mid contrasts, the resulting t-statistic is slightly stronger: in Hz, $t(14) = 6.4048, p < 0.001$; z-scored, $t(14) = 4.0197, p < 0.005$.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 13 of 29

Figure 3: Stressed vowel spaces for speakers ItF1, ItF3, ItF7, and ItF9. Each point represents word averages of F1, F2. Vowel quality (plotting symbol; SpeakerQ) assigned according to speaker judgments.

| Speaker	[e ɛ] distance		[o ɔ] distance		[front] – [back]	
	Hertz	**z-scored**	**Hertz**	**z-scored**	**Hertz**	**z-scored**
ItF1	496	1.314	159	0.792	337	0.522
ItF2	326	1.083	143	0.785	183	0.299
ItF3	572	1.529	222	1.027	350	0.502
ItF5	270	1.160	180	1.055	90	0.105
ItF6	211	0.688	56	0.351	155	0.337
ItF7	530	1.424	201	0.865	329	0.559
ItF8	53	0.226	30	0.163	22	0.063
ItF9	405	1.266	137	0.762	268	0.504
ItF10	831	2.128	198	1.041	633	1.087
ItF11	58	0.134	33	0.111	25	0.023
ItF12	396	0.987	146	0.920	250	0.067
ItF13	240	0.786	170	1.074	70	−0.288
ItF14	384	0.823	150	0.689	233	0.135
ItF15	462	1.362	195	1.000	267	0.362
ItM1	229	0.983	151	0.929	79	0.055
ItM2	293	1.345	132	0.881	161	0.464
ItM3	345	1.258	194	1.080	150	0.178

Table 4: Euclidean distance (Hertz vs. z-scored) between F1, F2 means of mid vowel pairs [e ɛ] and [o ɔ], as well as differences (Hertz vs. z-scored) between Euclidean distances in front and back vowels. Vowel quality assigned by speaker judgment.

space by a second metric that compares the distributions of adjacent pairs of vowels in order to quantify the amount of overlap among vowels in each speaker's data. The V1/ V2 Overlap measurement presented by Fougeron and Audibert (2011) was applied to F1 measurements of each speaker's stressed vowels. This formula is illustrated in (4) for the

pair /i e/; note that our calculations were applied to *z*-scored values rather than raw Hertz values.

(4) $i \: / \: e \: \: overlap = \left(\mu\left(F1_i\right) + \sigma\left(F1_i\right)\right) - \left(\mu\left(F1_e\right) - \sigma\left(F1_e\right)\right)$

For a contiguous vowel pair (e.g., /i e/), the standard deviation of the higher vowel (having lower F1) is added to its mean, whereas for the lower vowel the standard deviation is subtracted from the mean. This produces the upper limit of F1 standard deviation for the higher vowel, and the lower limit of F1 standard deviation for the lower vowel. Finally, the latter is subtracted from the former. Negative values indicate no overlap (that is, a positive distance in Hz) between the two vowels, while positive values express the amount of overlap in Hz between the two standard deviations. We computed overlap measurements for F1, among pairs of vowels that are adjacent in the front [i e, e ɛ] and back [ɔ o, o u]. **Figure 4** summarizes the amount of F1 overlap among each speaker's vowels, using Lobanov normalized values. If the reason that certain tokens are produced in the 'wrong' vowel cloud is overlap between the mid vowels, then we expect to see at least as much overlap between mid vowels as between the high-mid and high vowels. Instead, the results show that for 12 out of 17 speakers (or 12 out of 15, if ItF8 and ItF11 are excluded because of their lack of contrast), one of the mid vowel comparisons is that with the *least* amount of F1 overlap: For speakers ItF1, ItF2, ItF5, ItF7, ItF9, ItF10, ItF12, ItF15, ItM2 and ItM3, the [e ɛ] comparison has the least amount of overlap (expressed as *negative* bars
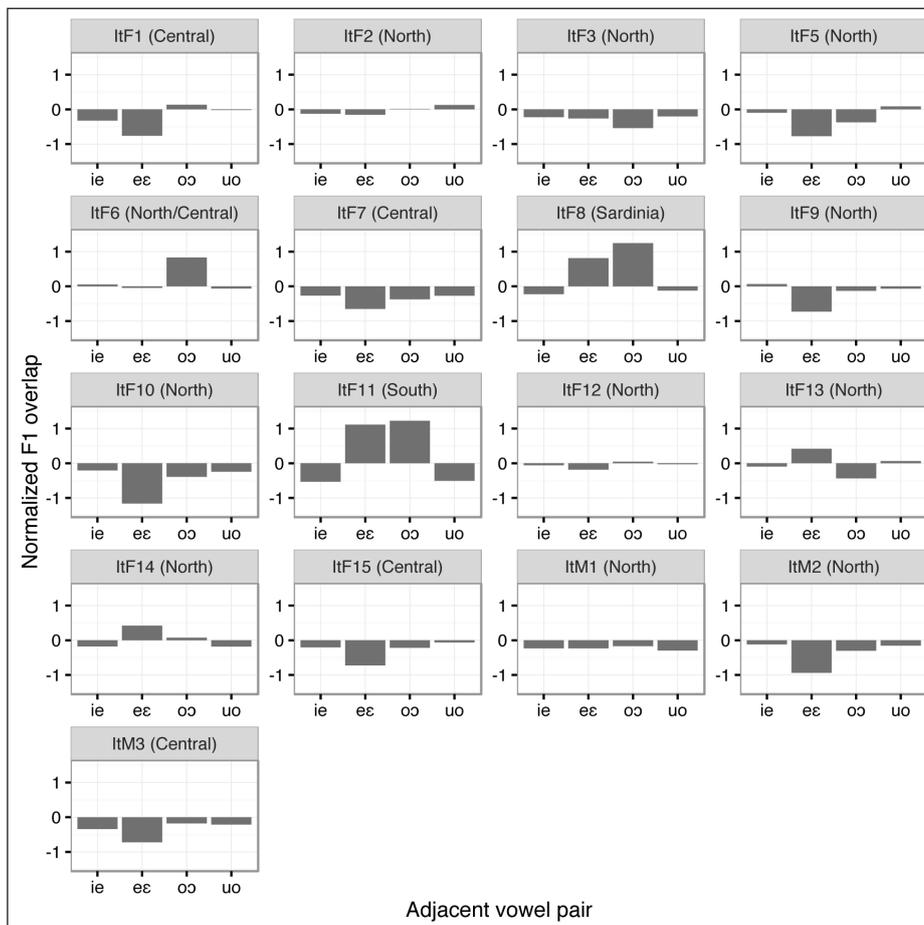
Figure 4: F1 overlap (in Lobanov normalized values) between contiguous front and back vowel pairs, per speaker. Mid vowel height judged by speaker. Negative values indicate separation.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 15 of 29

in the figure). For ItF3 and ItF13, the [o ɔ] comparison is the most distinct. This is the opposite of what we should expect if phonological closeness were entirely dependent on acoustic proximity or overlap, and further strengthens the conclusion that the sense of phonological closeness cannot be reduced to phonetic proximity in the vowel space.

Finally, the relationship between overlap and vowel height was tested in a two-way ANOVA with repeated measures in which overlap, calculated from normalized values, is the dependent variable. The model was restricted to data from the 15 speakers who have a demonstrable distinction between high-mid and low-mid vowels; ItF8 and ItF11 were excluded. The independent variables of Vowel Height (divided into high vs. high-mid and high-mid vs. low-mid) and Backness (divided into front and back), and their interaction, were within-subjects factors. The model returned a significant effect of Backness ($F(1, 14)$ $= 7.577, p < 0.05$), while Vowel Height ($F(1, 14) = 3.87, p = 0.0693$) and its interaction with Backness ($F(1, 14) = 3.289, p = 0.0912$) only approached a significance level of $p < 0.05$. Thus while there is a clear trend for greater acoustic height distance between [e ɛ] or [o ɔ] than between [i e] or [u o], it does not reach statistical significance in this sample.
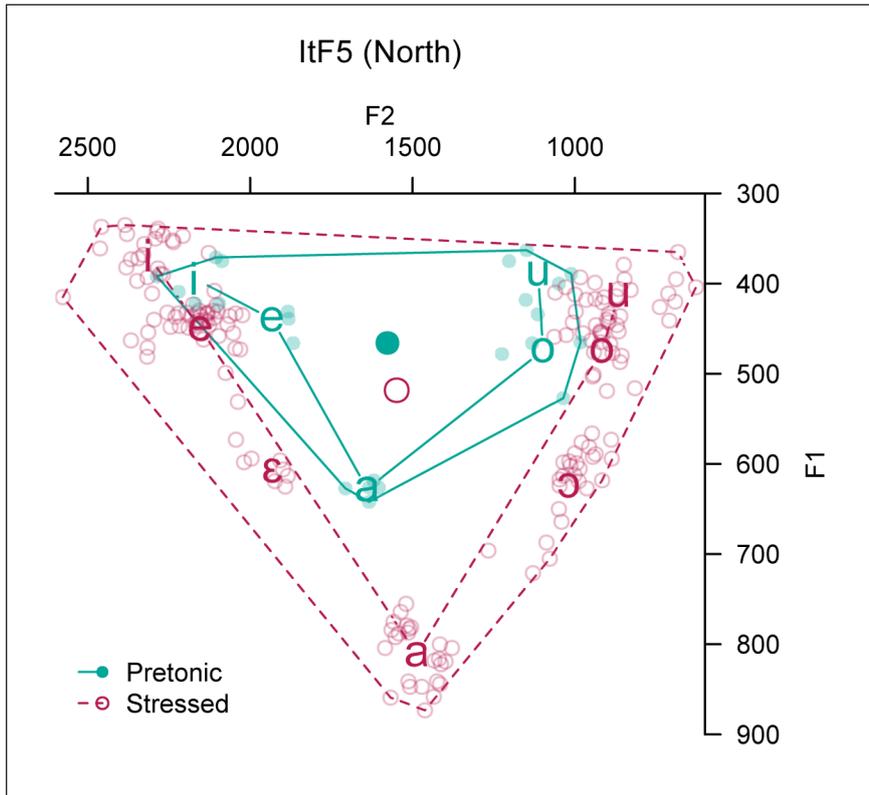
3.4 Phonologically conditioned effects on vowel quality

As noted in Section 2.1, different subsets of test words were intended to reveal the presence of phonological conditioning based on stress and syllable structure. Our findings on these potential conditioning factors are discussed here in turn. We note in advance that no effect of post-vocalic nasals was found in the DAN-da condition, and it is not further discussed.

3.4.1 Effects of stress

The data from pretonic target vowels show that for all speakers unstressed vowels are generally somewhat centralized. **Figures 5** and **6** illustrate this centralization for two speakers. Quantitatively, the degree of centralization across the vowel space was determined by calculating the Euclidean distance from each vowel centroid (i.e., mean F1, F2 values) to the overall vowel centroid (i.e., global mean F1, F2 values) for each speaker (Audibert et al., 2015; Fougeron & Audibert, 2011; Nadeu, 2013; Recasens & Espinosa, 2006). This distance is calculated for each vowel separately, and then averaged to arrive at a single value per speaker per stress condition. **Table 5**, which compares these average Euclidean distances across the stressed and pretonic conditions for all speakers, shows that stressed vowels have a greater average distance from the global centroid (i.e., are more peripheral) than pretonic vowels. It also shows the differences between the distances to the global centroid across stressed and pretonic vowels; these values are significantly different from zero (Hz: $t(16) = 11.369, p < 0.001$; z-scored: $t(16) = 13.026, p < 0.001$), indicating that the vowel space does shrink across stress conditions. A linear mixed-effects model, using Euclidean distances to speaker-specific centroids for all points, confirms varying effects of stress conditions across vowels; for details see Appendix 5.

While some centralization of unstressed vowels is expected, two specific aspects of the general finding are worth noting. First, in accordance with traditional descriptions of Standard Italian, mid vowel height distinctions are neutralized toward the higher category. This can be seen clearly in **Figures 5** and **6**: For both speakers shown in these figures, the pretonic mid vowels have F1 similar to that of stressed /e/ and /o/. The fact that this neutralization occurs in all speakers regardless of their regional background can be taken as further evidence of the validity of treating Standard Italian as an object of study. Second, pretonic /a/ is centralized to a striking degree, being realized on average with F1 values nearly 200 Hz lower than stressed /a/. This is somewhat surprising, since traditional descriptions of

Art. 19, page 16 of 29

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Figure 5: Pretonic vs. stressed vowel space for speaker ItF5. Each point represents a word average; outer convex hulls encompass all tokens, and interior lines connect individual vowel means. Overall centroids for pretonic vs. stressed conditions are also shown.
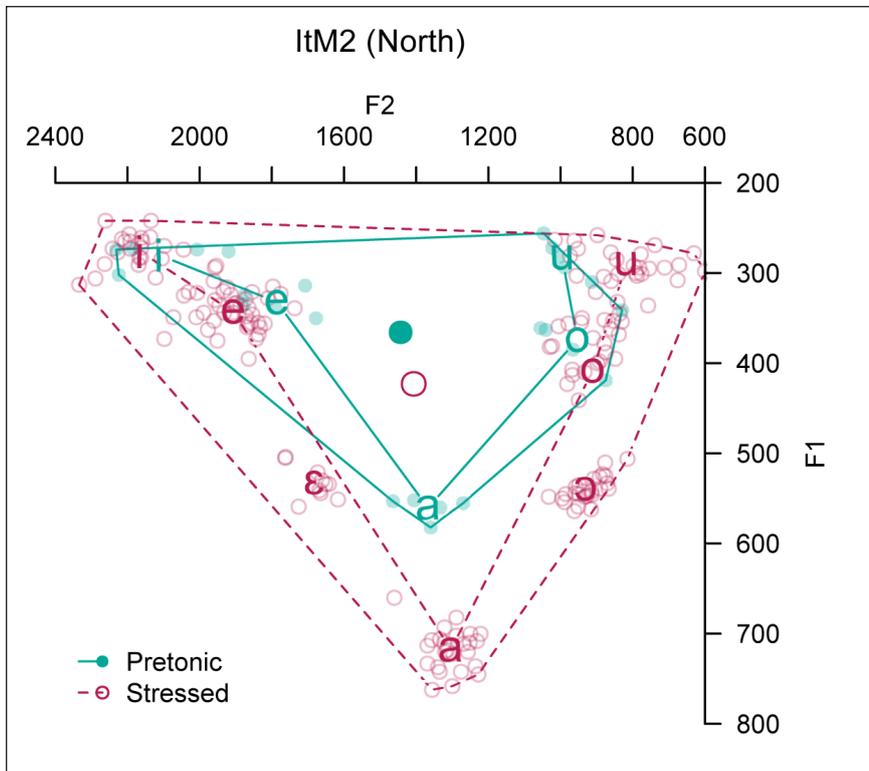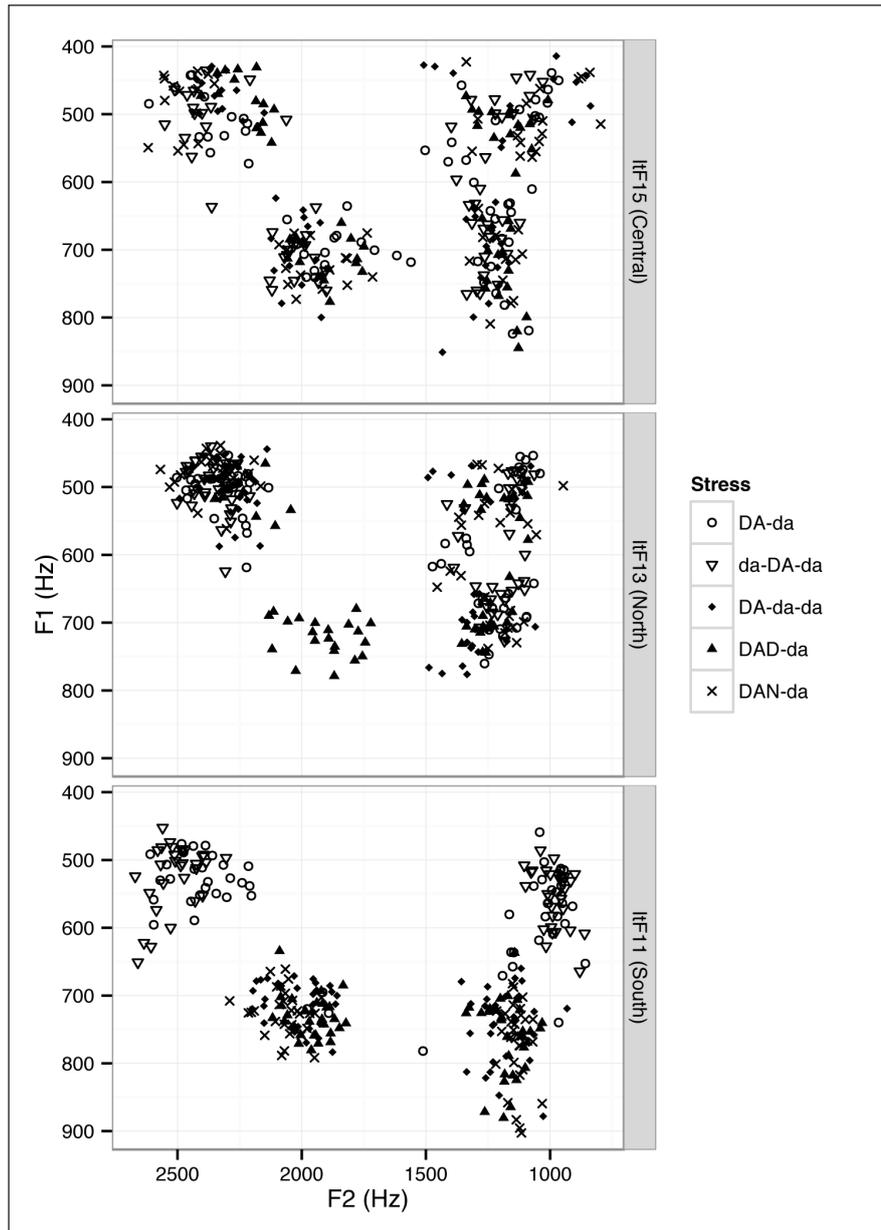

Figure 6: Pretonic vs. stressed vowel space for speaker ItM2. Each point represents a word average; outer convex hulls encompass all tokens, and interior lines connect individual vowel means. Overall centroids for pretonic vs. stressed conditions are also shown.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 17 of 29

Speaker	Stressed		Pretonic		[Stressed] – [Pretonic]	
	Hertz	z-scored	Hertz	z-scored	Hertz	z-scored
ItF1	656	1.3095	547	1.0430	109	0.2665
ItF2	632	1.3769	530	0.8529	176	0.5240
ItF3	683	1.2392	542	0.9555	148	0.2837
ItF5	561	1.2405	416	1.1115	48	0.1290
ItF6	697	1.3014	479	0.9845	146	0.3169
ItF7	702	1.2739	563	1.0209	127	0.2530
ItF8	573	1.3234	510	1.0654	92	0.2580
ItF9	671	1.2929	472	1.0452	102	0.2477
ItF10	641	1.3323	465	0.9783	141	0.3540
ItF11	650	1.3439	502	0.9410	145	0.4029
ItF12	586	1.3047	538	0.9287	218	0.3760
ItF13	548	1.3487	402	1.0301	139	0.3186
ItF14	688	1.2121	561	1.0614	63	0.1507
ItF15	581	1.3442	489	0.9241	199	0.4201
ItM1	549	1.3458	421	1.0266	128	0.3192
ItM2	494	1.3335	435	1.0843	59	0.2492
ItM3	486	1.3342	338	0.9467	148	0.3875

Table 5: Euclidean distance from vowel centroid (Hertz and z-scored), calculated as a per-speaker average with 7 vowels in stressed position, and 5 vowels in pretonic position, as well as differences (Hertz vs. z-scored) between Euclidean distances in stressed and pretonic vowels.

Standard Italian typically mention the absence of vowel reduction (Calamai, 2001; Savy & Cutugno, 1998). However, while the effect of stress on /a/ is numerically larger than for other vowels, as shown in Appendix 5, only for [o] is the difference in effect size significant. That is, all Italian vowels undergo centralization, though [o] is least affected.

3.4.2 Effects of syllable structure

As noted in the introduction, we investigated whether certain syllable structures privilege either high-mid or low-mid vowels. Such conditioning does occur, suggesting the regional variation we would expect from standard dialectological descriptions. The major effects we observe are illustrated in **Figure 7**, which shows data from stressed mid vowels for three speakers. In the vowel space of Central speaker ItF15, there is no relationship between the height of a mid vowel (in F1 space) and the phonological context in which it occurs (shown by plotting symbol). This speaker agrees almost entirely with prescriptive norms, so the phonological contexts are distributed evenly across the four mid vowel clouds in her speech. Northern speaker ItF13—like many of our Northern speakers—shows some evidence of phonological conditioning reminiscent of the French *loi de position*, with low-mid /ɛ/ limited to syllables with a non-nasal coda (DAD-da). However, the conditioning is by no means consistent: Some syllables of this type are realized with high-mid /e/, and among back vowels there is no apparent conditioning. By contrast, the Southern speaker ItF11 exhibits clear and consistent complementary distribution for both front and back vowels. Her high-mid vowels are found in stressed open penultimate syllables, while both closed syllables and antepenultimate open syllables contain low-mid vowels. Many Southern varieties show consistent allophonic distribution conditioned by syllable structure, though the details of the complementary distribution vary somewhat (Avolio, 1995); as

Art. 19, page 18 of 29
Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Figure 7: Vowel spaces for speakers ItF15, ItF13, and ItF11; stressed mid vowels only. Each point represents a single token.

seen in **Figure 2** above, this allophonic relationship between high-mid and low-mid vowels is accompanied by a lack of intuition about phonetic vowel height.

3.5 Interim summary: Judgments and the vowel space

The preceding subsections have provided initial answers to our research questions. First, we found considerable variation in speakers' judgments of mid vowel height across lexical items. At the same time, we also found that speakers are relatively accurate judges of their own productions. That is, the variable judgments are by and large an accurate reflection of speakers' actual pronunciation, meaning that many lexical items are variably pronounced. Some regional patterning was evident in this variation—in particular, Northern speakers tend to have [ɛ] in closed stressed syllables only—but otherwise, and especially for the back mid vowels, it seems valid to treat this lexical variability as a feature of Standard Italian and as a possible manifestation of phonological closeness.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 19 of 29

Second, we showed that the lexical variability cannot be explained in terms of phonetic closeness or overlap. Most speakers maintain seven phonetic vowel categories, and for most speakers, the mid vowel pairs [e ɛ] and [o ɔ] overlap phonetically *less* than [i e] or [o u]. This means that phonological closeness is not a matter of mere acoustic closeness or (at least for Northern and Central speakers) outright loss of mid vowel contrasts. We did, however, confirm that for all speakers the contrast is neutralized pretonically in favor of the higher mid vowels, which (as Trubetzkoy suggested for French) may also influence intuitions of closeness.

In the next section we document and discuss a further apparent manifestation of phonological closeness, namely a greater degree of mismatch between mid-vowel productions and speakers' judgments of their own usage.

4 Results II: Judgment – production mismatches
4.1 Identifying mismatches

In Section 3, having showed that speakers' judgments correspond to their productions better than to prescriptive vowel quality, we used speaker-assigned categories to plot and display data in different contexts. Despite the improved elliptical fit achieved by this method, there are still *mismatches*—cases in which speakers' judgments do not match their own pronunciation, so that the plotting symbol based on the speaker-assigned category shows up in the 'wrong' acoustic cloud. In this section we focus more precisely on the match between speakers' judgments and their productions. What we find broadly confirms that speakers are aware of their own usage, but it also makes clear that for a number of lexical items usage is actually variable *in individual speakers.*

To test for the prevalence of mismatches in a speaker's data, we must first identify genuine misclassifications, distinguishing them from those resulting from mere acoustic overlap. That is, in some cases, it is visually obvious that a given vowel token labeled as 'high-mid' is in the middle of a cloud of tokens labeled 'low-mid,' or vice-versa; but because the high-mid and low-mid clouds may overlap, it is possible that tokens appearing out of place in F1, F2 space were nevertheless correctly labeled by the speaker. We therefore use an objective measure to classify speakers' productions as either the higher or lower mid vowel. We do this by first grouping the acoustic data using a clustering algorithm. This is combined with a measure of acoustic distance, to determine which tokens are outliers with respect to the speaker's judgment. We evaluate all tokens that the clustering algorithm places in the 'wrong' cluster, to see if they are only peripheral to that cluster (and hence likely due to overlap) or near the heart of the cluster (in which case they are likely to be genuine mismatches).

All the analysis and discussion in this section excludes data from speakers ItF8 and ItF11, who as we showed in Section 3 do not have accurate intuitions about either their own speech or prescriptive norms, presumably owing to phonological merger in their regional varieties.

4.2 K-means clustering analysis

Acoustic data were grouped via k-means clustering. The k-means algorithm, applied in the two-dimensional vowel space, partitions data points in order to minimize the sum-of-squares distance between the points and their assigned cluster. The analysis was applied to speaker-normalized F1, F2 data, and it was conducted in R using the `kmeans()` function, applying the Hartigan and Wong (1979) algorithm with 100 random starts. It was run separately for each adjacent pair of stressed vowels, for each speaker. It was necessary to use normalized (z-scored) data because raw F1 and F2 values fall along scales of different magnitude, which the k-means algorithm does not interpret accurately; likewise, the

algorithm was restricted to two vowels at a time in order to avoid grouping front and back vowels in the same cluster. The algorithm was directed to divide data into two clusters for each of six comparison pairs: [i e, e ɛ, ɛ a, a ɔ, ɔ o, o u], with mid vowel targets placed into comparisons according to each speaker's judgment.

The output of the *k*-means clustering is a list of cluster assignments for each token. **Figure 8** shows an example, using speaker ItF9's mid vowels only: On the left, individual tokens are labeled according to her judgments, while on the right they are labeled with the cluster assigned by *k*-means. Although the outcomes for front mid vowels are identical across the two labeling techniques, *k*-means results in less overlap among the back vowel ellipses. Tokens that are labeled differently between the two plots are potential mismatches; note that some of these occur far from the boundary between high-mid and low-mid vowels, for example in tokens judged [o] that have an F1 of more than 650 Hz (visible in the SpeakerQ panel).

4.3 Identifying acoustic overlap

All mid vowel tokens whose *k*-means cluster label did not match SpeakerQ are candidates for judgment-production mismatches. However, because vowel clouds overlap in F1, F2 space, it is necessary to identify the subset of tokens that are also acoustically distant from the quality assigned by SpeakerQ. We therefore combine the *k*-means clustering results with a second method of classifying vowels, to identify outliers based on their distance from the center of each vowel cloud. We calculated the Mahalanobis distance (Mahalanobis, 1936) between each token and the central formant tendency. This measure is similar to Euclidean distance, but it takes into account the possibility of an ellipsoidal cloud of data rather than a spherical one. It is the unitless, directionless distance of each data point from the center of the cloud to which it belongs, divided by the width of the ellipsoid as calculated in the direction of that data point, and can be considered a multidimensional *z*-score (Labov et al., 2013).

Mahalanobis distances were calculated using *z*-scored F1, F2 values for each speaker's stressed vowel tokens, labeled according to speaker judgments. Calculations were carried out using the `cov.trob()` command to estimate covariance for a multivariate *t*-distribution (Kent et al., 1994; Venables & Ripley, 2002), in conjunction with the `mahalanobis()` command in R's *stats* package (R Core Team, 2000). For each adjacent-vowel
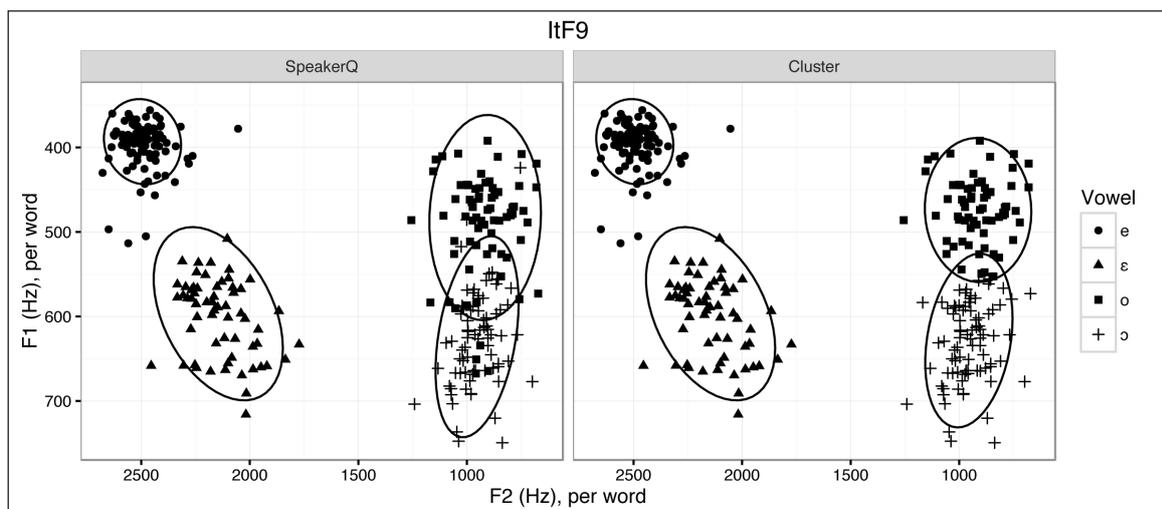

Figure 8: Vowel space for speaker ItF9; individual stressed mid vowel tokens plotted in raw Hertz. Left: Vowel quality (plotting symbol) assigned according to speaker judgments. Right: Vowel quality assigned according to k-means clustering. Ellipses represent 95% confidence intervals.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 21 of 29

comparison [i e, e ɛ, ɛ a, a ɔ, ɔ o, o u], we then identified all vowel tokens for which *two* conditions were met: First, the token was assigned by *k*-means to the *opposite* cluster from that identified by the speaker; second, the token's Mahalanobis distance classified it as an outlier, placing it outside the 95% quantile of a Chi-Squared distribution with two degrees of freedom. Together, the two conditions (*k*-means mismatch plus large Mahalanobis distance) highlight *categorical mismatches,* in which the vowel token is realized squarely in the opposite cloud to that judged by the speaker. The result of this conservative analysis technique is shown for two speakers in **Figure 9**, in which the 'mismatch' label is applied. These tokens are generally restricted to the center of 'opposite' vowel clouds (e.g., the center of the [ɛ] cloud for ItF12, or the [ɔ] cloud for ItF7).

4.4 Speaker misclassifications of mid vowel height

The number of categorical mismatches for each speaker, per vowel pair, is summarized in **Table 6**. It contains mismatches for all vowel pairs, each pair comprising between 2,000 and 3,000 data points across all speakers. The ubiquity of mismatches across comparisons suggests that even the combined method just described is affected by acoustic overlap in the vowel space, since the mismatches involving the corner vowels [i a u] do not represent misjudgment by the speaker and must be regarded as artifacts of the algorithmic classification. This interpretation is strengthened by the greater number of mismatches among back vowels than among front vowels, since greater acoustic overlap is expected to result from the relatively reduced F1 range available in that portion of the vowel space (De Boer, 2011). However, it is also clear that many more mismatches are identified for the mid vowel pairs ([e ɛ] and [ɔ o]) than for the pairs involving [i a u]. Even if all pairs yield some spurious mismatches due only to acoustic overlap, the excess in the two mid vowel pairs can be assumed to include genuine mismatches between speakers' judgments and their own productions.

To confirm that the differences between the columns of **Table 6** are genuine, we tested the relationship between categorical mismatches and vowel height in a two-way ANOVA with repeated measures. The number of categorical mismatches was the dependent variable, and the independent variables of Adjacent Vowel Comparison (divided into high vs. high-mid, high-mid vs. low-mid, and low-mid vs. low) and Backness (front vs. back) were treated as within-subjects factors. The model returned significant effects of height
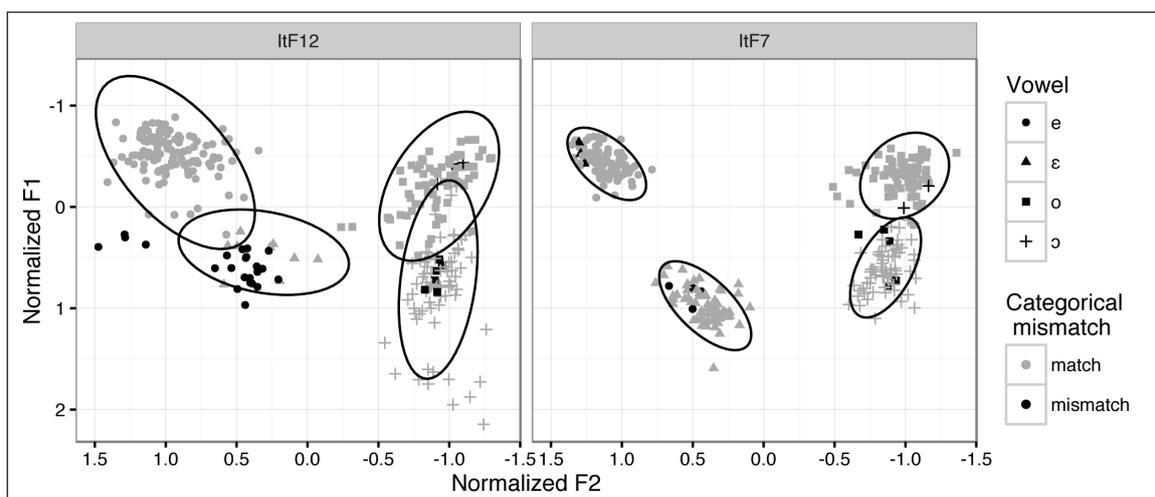
Figure 9: Mid vowels of speakers ItF12 and ItF7, labeled according to the categorical mismatch criterion (*k*-means and Mahalanobis distance); mismatches are black; 95% confidence ellipse based on speaker judgment.

Speaker	[i e]	[e ɛ]	[ɛ a]	[a ɔ]	[ɔ o]	[o u]
ItF1	1	0	2	3	13	3
ItF2	0	16	2	1	10	10
ItF3	0	9	5	5	5	11
ItF5	2	6	1	6	8	4
ItF6	1	12	0	5	10	3
ItF7	0	7	2	1	7	3
ItF9	3	0	0	0	12	4
ItF10	2	3	7	8	8	4
ItF12	4	24	0	9	9	2
ItF13	5	4	0	2	10	3
ItF14	1	13	0	1	7	4
ItF15	3	2	0	7	8	6
ItM1	3	6	5	0	13	2
ItM2	7	0	1	0	10	3
ItM3	4	1	2	8	11	1
Total	36	103	27	56	141	63

Table 6: Number of categorical mismatches per speaker, per adjacent vowel comparison. A categorical mismatch indicates that the cluster assigned by k-means does not match the speaker's judgment (or orthographic representation, for [i a u]), and that the vowel lies outside the 95% confidence ellipse for Mahalanobis distance.

comparison ($F(2, 28) = 22.62, p < 0.001$) and backness ($F(1, 14) = 13.24, p < 0.005$). A post-hoc pairwise t-test with Bonferroni correction for multiple comparisons showed that the number of categorical mismatches differed significantly between the high-mid vs. low-mid ([e ɛ] and [o ɔ]) comparisons and high vs. high-mid ([i e] and [u o]) comparisons ($p < 0.001$), and between the high-mid vs. low-mid and low-mid vs. low ([ɛ a] and [a ɔ]) comparisons ($p < 0.001$); however, the number of mismatches across low-mid vs. low and high vs. high-mid comparisons was not significantly different ($p = 1$). In other words, the number of categorical mismatches between high-mid and low-mid vowels, both front and back, is significantly larger than those found for other vowel comparisons.

It is worth noting that in **Table 6**, there may be a tendency towards more mismatches between [e ɛ] for Northern speakers (e.g., ItF2, ItF12, ItF14). This could indicate decreased speaker awareness of the front mid vowel contrast for these speakers; for example, a preponderance of ItF12's [e ɛ] mismatches occurs in closed target syllables, where phonological conditioning is present. However, regardless of speakers' region, mismatches occur between [o] and [ɔ] as well. This indicates that even for Central speakers, who appear to have the clearest intuitions and most closely match the prescriptive standard, these judgment-production mismatches do occur.

4.5 Interpreting mid-vowel mismatches

The foregoing analysis confirms that there are significantly more mismatches in mid-vowel pairs than in pairs consisting of a mid vowel and a corner vowel. Even if all the cases involving the corner vowels are dismissed as spurious, a proportion of those involving mid vowels are expected to represent genuine mismatches. However, such mismatches do not necessarily involve phonetic confusion. Only in a few cases are there consistent misjudgments, where a speaker reports using vowel A in a given word but in fact uses vowel B in all three repetitions of that word; the average number of mid-vowel words misjudged in

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 23 of 29

this way is only 2.4 per speaker, out of the 100 mid-vowel words in the full set of materials (median 2, range 0–8). Our conclusion in Section 3 that speakers are broadly aware of their own productions appears well justified.

Instead, most of the mid vowel mismatches in **Table 6** involve cases of *variable production*, where (irrespective of speakers' judgments) one repetition of a given word contains vowel A and the other two contain vowel B. Since speakers were not making phonetic judgments about their recorded productions but were providing one metalinguistic judgment for each of the 100 test words, any variability across repetitions will necessarily lead our analysis to detect mismatches. On average, speakers exhibit such variability in 7.2 of the 100 test words (median 6, range 3–11). Some of these cases, of course, may be spurious, due to acoustic overlap among vowels, but it is likely that many are genuine. That is, the choice of mid vowel in some words varies not only regionally, but also in individual speakers. This intraspeaker variability occurred in approximately 10% of productions; on average, for 90 of the 100 mid-vowel words in our materials (median 91, range 86–95) speakers used the same vowel in all three repetitions and accurately reported its quality. Within that range, speakers from the North and Central regions performed equivalently.

In summary, alongside the clear patterns of broadly accurate speaker judgments of mid vowel height shown in Section 3, there are also mismatches, where speakers' judgment of the vowel quality in their own pronunciation of a given word is at odds in some way with their actual production. Whether these mismatches involve real speaker uncertainty or the effect of underspecified or variable lexical representations, they set /e ɛ ɔ o/ apart from /i a u/ on a dimension of phonological contrast, and are a likely component of phonological closeness.

5 Discussion and Conclusion

Our study shows that the mid-vowel pairs in Standard Italian, at least as spoken in Central and Northern regions of Italy, are phonetically distinct, and that speakers generally have good intuitions about which vowel they use in which word. On the face of it, this could be taken to suggest that our original motivation—understanding the nature of the phonological closeness described for Italian—was simply misplaced, and that the Italian mid vowels pose no sort of theoretical conundrum for a traditional notion of phonemic contrast. With regard to Kiparsky's 2x2 matrix pitting contrastiveness against distinctiveness, we find that for nearly all speakers, the high-mid and low-mid vowels are distinctive, because they have seven distinct vowel clouds. The existence of minimal pairs, and the lack (for Central speakers) of any phonological conditioning, indicates that at least for some speakers, Italian mid vowels are also contrastive and could therefore be considered separate phonemes.

However, we have documented the existence of significant lexical variability between speakers in the choice of mid vowel in stressed syllables, and we have shown that all speakers' productions do not always match their judgment about which vowel they use in which word. This is true even though the acoustic overlap between mid vowels tends to be less than that for other adjacent vowels, and the mid vowels themselves are far apart in Euclidean distance. In other words, *despite* the presence of both minimal pairs and a robust phonetic distinction, speakers' lexical intuitions about the Italian mid vowel distinctions are weaker than we might expect. This finding is comparable to the differences in identifiability of /ɨ/ and /ʌ/ in Romanian documented by Renwick (2014). It appears that Trubetzkoy's puzzling 'particular closeness' is a real phenomenon.

What do the Italian mid vowels tell us about the essential characteristics of this closeness in phonology generally? Except for the Sardinian speaker, it is certainly not a matter of any loss of phonetic distinctiveness: The mid vowel categories are acoustically separate and if

anything, overlap less than other pairs of Italian vowels. Neither of Kiparsky's intermediate cases from **Table 1** (near contrast and allophonic awareness) applies; all our Northern and Central speakers are clearly aware of the phonetic distinction, and especially for Central speakers, we are certainly not dealing with phonologically conditioned allophones. Nor can we unreservedly accept Trubetzkoy's original speculation about the source of the particular closeness, namely the neutralization of the mid vowel contrasts in pretonic position. While this may make some contribution to the sense of closeness, it cannot by itself constitute the explanation; as Ladd (2006) observed, there are many cases of neutralization (e.g., German final devoicing) that do not give rise to a feeling of closeness between phonemes. Instead, we propose that closeness resides somewhere in the relationship between phonetic categories and lexical use, or the mapping of phonetic categories to specific words. The phonetic categories—which yield Kiparsky's distinctiveness—are sharply defined; the use to which the distinction is put in the language—Kiparsky's contrastiveness—is both variable and limited. Phonological closeness arises because *contrastiveness can be a matter of degree.*[6]

Obvious ways in which the Italian mid vowel distinctions are less contrastive than other phonemic contrasts include functional load (i.e., count of minimal pairs) and distributional predictability (i.e., any effects of phonological conditioning). On the first of these measures, the mid vowel distinctions in Italian are clearly 'not very contrastive;' on the second, we have confirmed that Northern speakers tend to show a good deal of phonological predictability among front mid vowels. However, we propose that a more important aspect is variability in individual lexical items. Our results demonstrate the existence of lexical variation across speakers, and show that regional effects drive a considerable amount of this variation. We suggest that this variability affects the phonological representations of speakers of Standard Italian. To the extent that they hear a mixture of regional varieties in their daily lives (e.g., in large cities and via national broadcast media), many speakers will have the experience of ignoring mid vowel quality in making lexical identifications during normal speech comprehension. Although the limited functional load of the phonetic distinction and the general redundancy of the speech signal ensure that misunderstandings will almost never arise, this variation could contribute to a weakening of their phonological representations and intuitions, as has recently been suggested for Catalan (Nadeu & Renwick, 2016).

It may also be relevant that some variability in the vowel quality of individual lexical items results not from individual or regional variation, but from the pretonic neutralization of mid-vowel quality, whose existence we have also clearly demonstrated. Although, as just noted, we find it unlikely that neutralization *per se* results in phonological closeness, this particular neutralization process results in vowel alternations in many lexical stems. For example, virtually all verbs with mid-vowels in the stem-final syllable are potentially affected (e.g., *io ˈfreno* 'I brake' but *freˈnare* 'to brake'). Given the liberal use of diminutive suffixes and Italian's generally rich derivational morphology, the mid vowels in many noun and adjective stems are affected in the same way (e.g., *ˈsonno* 'sleep' but *sonnelˈlino* 'nap, snooze'; *ˈcentro* 'center' but *cenˈtrale* 'central'). This variation may further destabilize lexical vowel identity.[7]

---

[6] Throughout this discussion we focus on the phonological implications of particular closeness in general, and refrain from any speculation about the historical sources of the phenomenon for Italian mid vowels. It is likely, as suggested by an anonymous reviewer, that one contributory factor to the Italian case is the absence of an orthographic distinction, and that interference between Standard Italian and regional dialects may also be relevant. But we emphasize that trying to explain the source of a specific instance of phonological closeness may be as ultimately futile as trying to explain the source of a specific phonemic contrast. Part of the motivation for our investigation is the contention that the classical dichotomy between phoneme and allophone is an oversimplification, and that we need to move beyond it.

[7] Word frequency may also be relevant: Less-frequent lexical items, such as those created by affixation, might have greater lexical instability than frequent words, a hypothesis supported by recent results on New Zealand English (Hay et al., 2015) in which low-frequency words lead sound change.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts

Art. 19 page 25 of 29

In summary, while our results clearly seem to show the usefulness of Kiparsky's distinction between distinctiveness and contrastiveness, we believe that a fuller understanding of the various kinds of intermediate phonological relationships catalogued by Goldsmith (1995) or Hall (2013) will require us to recognize that the function of phonetic distinctions cannot be reduced to the presence or absence of contrast. We suggest that recent attempts to define a quantitative notion of 'degree of contrastiveness' are more likely than those built on purely categorical differences to lead to important theoretical progress in providing for a richer taxonomy of qualitative distinctions. To advance this line of inquiry, it would be valuable to compare metalinguistic judgments and productions of uncontroversially contrastive phonemic distinctions with those for putative cases of phonological closeness and near contrast in a variety of languages, to establish some kind of baseline for evaluating behavioral results. Such efforts are already underway for other Romance languages, in recent studies by Hall and Hume (2015) on French and Nadeu and Renwick (2016) on Catalan. We hope that the present study, by showing how categorical phonetic distinctiveness can coexist with limited lexical contrastiveness, will contribute to this overall line of research.

Additional Files

The additional files for this article can be found as follows:

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Abercrombie, D. 1979. The accents of Standard English in Scotland. In: A. J. Aitken and T. McArthur eds., *Languages of Scotland* (pp. 68–84). Edinburgh: W. & R. Chambers.

Adank, P., Smits, R., and Van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America, 116*(5), 3099–3107. DOI: http://dx.doi.org/10.1121/1.1795335

Audibert, N., Fougeron, C., Gendrot, C., and Adda-Decker, M. 2015. Duration- vs. style-dependent vowel variation: A multiparametric investigation. In *Proceedings of the 18th*

International Congress of Phonetic Sciences. Glasgow, Scotland. Retrieved from http://
www.icphs2015.info/pdfs/Papers/ICPHS0753.pdf

Avolio, F. 1995. *Bommèspre: profilo linguistico dell'Italia centro-meridionale*. San Severo,
Italy: Gerni.

Bates, D. and Maechler, M. 2009. lme4: Linear mixed-effects models using S4 classes (Ver-
sion R package version 0.999375–32).

Bertinetto, P. M. and Loporcaro, M. 2005. The sound pattern of Standard Italian, as com-
pared with the varieties spoken in Florence, Milan and Rome. *Journal of the Inter-
national Phonetic Association, 35*(2), 131–151, DOI: http://dx.doi.org/10.1017/
S0025100305002148

Bigi, B. and Hirst, D. 2012. Speech Phonetization Alignment and Syllabification (SPPAS):
a tool for the automatic analysis of speech prosody. In *Speech Prosody* (pp. 19–22).
Shanghai (China): Tongji University Press. Retrieved from http://www.sppas.org/

Bloch, B. 1941. Phonemic overlapping. *American Speech, 16*(4), 278–284. DOI: http://
dx.doi.org/10.2307/486567

Boersma, P. and Weenink, D. 2015. Praat: Doing phonetics by computer [Computer pro-
gram], Version 5.4.08 (Version 5.4.08). Retrieved from http://www.praat.org

Bolognesi, R. 1998. The phonology of Campidanian Sardinian (Doctoral dissertation).
University of Amsterdam (Holland Institute of Generative Linguistics: HIL Disserta-
tions, no. 38).

Calamai, S. 2001. Il vocalismo atono della varietà pisana. Prime evidenze sperimentali.
Quaderni del Laboratorio di Linguistica, 2, 64–82.

Canepari, L. 1979. *Introduzione alla fonetica*. Torino: G. Einaudi.

Canepari, L. 1980. *Italiano standard e pronunce regionali*. Padua: CLUEP.

Canepari, L. 1999. *Manuale di pronuncia italiana*. Bologna: Zanichelli.

Chen, Y. 2008. The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonet-
ics, 36*(4), 629–648. DOI: http://dx.doi.org/10.1016/j.wocn.2008.03.001

Chitoran, I. 2002. A perception-production study of Romanian diphthongs and glide-
vowel sequences. *Journal of the International Phonetic Association, 32*(2), 203–222. DOI:
http://dx.doi.org/10.1017/S0025100302001044

De Boer, B. 2011. First formant difference for /i/ and /u/: A cross-linguistic study and an
explanation. *Journal of Phonetics, 39*(1), 110–114. DOI: http://dx.doi.org/10.1016/j.
wocn.2010.12.005

De Mauro, T. 1976. *Storia linguistica dell'Italia unita*. Roma; Bari: Laterza.

De Mauro, T. 2000. *Il dizionario della lingua italiana*. Torino: Paravia.

Dinnsen, D. A. and Charles-Luce, J. 1984. Phonological neutralization, phonetic imple-
mentation and individual differences. *Journal of Phonetics, 12*(1), 49–60.

Ernestus, M. and Baayen, R. H. 2006. The functionality of incomplete neutralization in
Dutch: The case of past-tense formation. In: L. Goldstein, D. H. Whalen and C. Best eds.,
Laboratory Phonology, 8 (pp. 27–49). Berlin; New York: Mouton de Gruyter.

Fagyal, Z., Kibbee, D. and Jenkins, F. 2006. *French: A linguistic introduction*. Cambridge
University Press. DOI: http://dx.doi.org/10.1017/cbo9780511791185

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological
Bulletin, 76*(5), 378–382. DOI: http://dx.doi.org/10.1037/h0031619

Fougeron, C. and Audibert, N. 2011. Testing various metrics for the description of vowel
distortion in dysarthria. In *Proceedings of the 17th International Congress of Phonetic Sci-
ences* (pp. 1–4). Hong Kong.

Fries, C. C. and Pike, K. L. 1949. Coexistent phonemic systems. *Language, 25*(1), 29–50.
DOI: http://dx.doi.org/10.2307/409907

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 27 of 29

Gabrielli, A. 1956. *Dizionario linguistico moderno; guida pratica per scrivere e parlar bene.* Milan: Edizioni Scolastiche Mondadori.

Gamer, M., Lemon, J., Fellows, I., and Singh, P. 2012. Irr: Various Coefficients of Interrater Reliability and Agreement. *R Package Version 0.84*.

Goldsmith, J. A. 1995. Phonological Theory. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* (pp. 1–23). Cambridge, MA: Blackwell Publishers.

Hajek, J. and Maeda, S. 2000. Vowel height and duration on the development of distinctive nasalization. In: M. B. Broe and J. B. Pierrehumbert eds., *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 52–69). Cambridge, UK: Cambridge University Press.

Hall, K. C. 2009. *A Probabilistic Model of Phonological Relationships from Contrast to Allophony* (Doctoral dissertation). The Ohio State University, Columbus, OH.

Hall, K. C. 2012. Phonological relationships: A probabilistic model. *McGill Working Papers in Linguistics*, *22*(1), 1–14.

Hall, K. C. 2013. A typology of intermediate phonological relationships. *The Linguistic Review*, *30*(2), 215–275. DOI: http://dx.doi.org/10.1515/tlr-2013-0008

Hall, K. C. and Hume, E. 2015, August. *Modeling Perceived Similarity: The influence of phonetics, phonology and frequency on the perception of French vowels*. Manuscript, submitted August 2015.

Hartigan, J. A. and Wong, M. A. 1979. A *k*-means clustering algorithm. *Applied Statistics*, *28*, 100–108. DOI: http://dx.doi.org/10.2307/2346830

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., and LaShell, P. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91. DOI: http://dx.doi.org/10.1016/j.cognition.2015.02.012

Hockett, C. F. 1966. *The quantification of functional load: A linguistic problem* (U.S. Air Force Memorandum No. RM-5168-PR).

Hualde, J. I. and Prieto, M. 2002. On the diphthong/hiatus contrast in Spanish: some experimental results. *Linguistics*, *40*(2), 217–234. DOI: http://dx.doi.org/10.1515/ling.2002.010

Jones, M. 1988. Sardinian. In: M. Harris and N. Vincent eds., *The Romance Languages* (pp. 314–350). New York: Oxford University Press.

Kent, J. T., Tyler, D. E., and Vardi, Y. 1994. A curious likelihood identity for the multivariate t-distribution. *Communication in Statistics – Simulation and Computation*, *23*, 441–453. DOI: http://dx.doi.org/10.1080/03610919408813180

Kiparsky, P. 2014. New perspectives in historical linguistics. In: C. Bowern and B. Evans eds., *The Routledge Handbook of Historical Linguistics* (pp. 64–102). London, New York: Routledge.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R Package Version*, 2–0.

Labov, W. 1989. Exact description of the speech community: Short A in Philadelphia. In: R. Fasold and D. Schiffrin eds., *Language change and variation* (pp. 1–57). Washington, D.C.: Georgetown University Press. DOI: http://dx.doi.org/10.1075/cilt.52.02lab

Labov, W. 1994. *Principles of language change: Internal factors*. Oxford: Blackwell.

Labov, W., Rosenfelder, I., and Fruehwald, J. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*(1), 30–65. DOI: http://dx.doi.org/10.1353/lan.2013.0015

Labov, W., Yaeger, M., and Steiner, R. 1972. *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.

Ladd, D. R. 2006. "Distinctive phones" in surface representation. In: L. Goldstein, D. H. Whalen, and C. T. Best eds., *Laboratory Phonology, 8* (pp. 3–26). Berlin; New York: Mouton de Gruyter.

Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. DOI: http://dx.doi.org/10.2307/2529310

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America, 49*(2B), 606–608. DOI: http://dx.doi.org/10.1121/1.1912396

Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta), 2*, 49–55.

Moulton, W. G. 1947. Juncture in modern standard German. *Language, 23*(3), 212–226. DOI: http://dx.doi.org/10.2307/409876

Nadeu, M. 2013. *The effects of lexical stress, intonational pitch accent and speech rate on vowel quality in Catalan and Spanish* (Doctoral dissertation). University of Illinois at Urbana-Champaign.

Nadeu, M. 2014. Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics, 46*, 1–22, DOI: http://dx.doi.org/10.1016/j.wocn.2014.05.003

Nadeu, M. and Renwick, M. E. L. 2016. Variation in the lexical distribution and implementation of phonetically similar phonemes in Catalan. *Journal of Phonetics, 58*, 22–47, DOI: http://dx.doi.org/10.1016/j.wocn.2016.05.003

Pittau, M. 1991. *Grammatica della Lingua Sarda (Varietà Logudorese)*. Sassari: Carlo Delfino editore.

Port, R. and Crawford, P. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*(4), 257–282.

R Core Team. 2000. *R Language Definition*. Available from CRAN sites. Retrieved from ftp://155.232.191.133/cran/doc/manuals/r-devel/R-lang.pdf

Rebora, P. 1958. *Cassell's Italian-English, English-Italian dictionary*. London: Cassell.

Recasens, D. 1990. The articulatory characteristics of palatal consonants. *Journal of Phonetics, 18*(2), 267–280.

Recasens, D. and Espinosa, A. 2006. Dispersion and variability of Catalan vowels. *Speech Communication, 48*(6), 645–666. DOI: http://dx.doi.org/10.1016/j.specom.2005.09.011

Recasens, D., Farnetani, E., and Ní Chasaide, A. 1991. A study of the production and coarticulatory characteristics of palatal and palatalized consonants. *ACCOR Periodic Progress Report, 2*.

Recasens, D., Fontdevila, J., and Pallarès, M. D. 1995. A production and perceptual account of palatalization. In: B. Connell and A. Arvaniti eds., *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 265–281). Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511554315.019

Renwick, M. E. L. 2014. *The Phonetics and phonology of contrast: The case of the Romanian vowel system*. Berlin, Boston: De Gruyter Mouton. DOI: http://dx.doi.org/10.1515/9783110362770

Rogers, D. and d'Arcangeli, L. 2004. Italian. *Journal of the International Phonetic Association, 34*(1), 117–121, DOI: https:/doi.org/10.1017/S0025100304001628

Rohlfs, G. 1972. *Studi e ricerche su lingua e dialetti d'Italia*. Firenze: G. C. Sansoni.

Savino, M. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association, 42*(1), 23–48, DOI: http://dx.doi.org/10.1017/S002510031100048X

Savy, R. and Cutugno, F. 1998. Hypospeech, vowel reduction, centralization: how do they interact in diaphasic variations. In *Proceedings of the XVIth International Congress of Linguists*. Oxford, UK: Pergamon.

Renwick and Ladd: Phonetic Distinctiveness vs. Lexical Contrastiveness in
Non-Robust Phonemic Contrasts

Art. 19 page 29 of 29

Scherrer, Y., Boula de Mareüil, P., and Goldman, J.-P. 2015. Crowdsourced mapping of pronunciation variants in European French. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland. Retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0421.pdf

Scobbie, J. M. 2005. The phonetics phonology overlap. *QMU Speech Science Research Centre Working Papers, WP-1*.

Scobbie, J. M. and Stuart-Smith, J. 2008. Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In: P. Avery, B. E. Dresher, and K. Rice eds., *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 87–114). Berlin: de Gruyter.

Slowiaczek, L. M. and Dinnsen, D. A. 1985. On the neutralizing status of Polish word final devoicing. *Journal of Phonetics, 13*(3), 325–341.

Trubetzkoy, N. S. 1936. Die Aufhebung der phonologischen Gegensätze. *Travaux Du Cercle Linguistique de Prague, 6*, 29–45.

Trubetzkoy, N. S. 1939 [1969]. *Principles of Phonology*. (C. Baltaxe, trans.). Berkeley: University of California Press.

Venables, W. N. and Ripley, B. D. 2002. *Modern Applied Statistics with S* (4th ed.). New York: Springer. DOI: http://dx.doi.org/10.1007/978-0-387-21706-2

Vincent, N. 1988. Italian. In: M. Harris & N. Vincent eds., *The Romance Languages* (pp. 279–313). New York: Oxford University Press.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*(2), 251–276, DOI: http://dx.doi.org/10.1016/S0095-4470(03)00032-9

Wedel, A., Jackson, S., and Kaplan, A. 2013. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech, 56*(3), 395–417. DOI: http://dx.doi.org/10.1177/0023830913489096

Wieling, M., Montemagni, S., Nerbonne, J., and Baayen, R. H. 2014. Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language, 90*(3), 669–692. DOI: http://dx.doi.org/10.1353/lan.2014.0064