ELSEVIER

Speech Communication 20 (1996) 255–272

SPEECH COMMUNICATION

Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics [1,2]

Ann R. Bradlow *, Gina M. Torretta, David B. Pisoni

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405, USA

Received 27 March 1996; revised 8 August 1996

Abstract

This study used a multi-talker database containing intelligibility scores for 2000 sentences (20 talkers, 100 sentences), to identify talker-related correlates of speech intelligibility. We first investigated "global" talker characteristics (e.g., gender, F0 and speaking rate). Findings showed female talkers to be more intelligible as a group than male talkers. Additionally, we found a tendency for F0 range to correlate positively with higher speech intelligibility scores. However, F0 mean and speaking rate did not correlate with intelligibility. We then examined several fine-grained acoustic-phonetic talker-characteristics as correlates of overall intelligibility. We found that talkers with larger vowel spaces were generally more intelligible than talkers with reduced spaces. In investigating two cases of consistent listener errors (segment deletion and syllable affiliation), we found that these perceptual errors could be traced directly to detailed timing characteristics in the speech signal. Results suggest that a substantial portion of variability in normal speech intelligibility is traceable to specific acoustic-phonetic characteristics of the talker. Knowledge about these factors may be valuable for improving speech synthesis and recognition strategies, and for special populations (e.g., the hearing-impaired and second-language learners) who are particularly sensitive to intelligibility differences among talkers.

Zusammenfassung

Auf der Grundlage eines Multi-Sprecher Datenkorpus mit den Verständlichkeitsbewertungen von 2000 Sätzen (20 Sprecher, 100 Sätze) wurden in dieser Studie die sprecherspezifischen Korrelate der Sprachverständlichkeit untersucht. Zunächst wurden "globale" Sprechercharakteristika untersucht (z.B. Geschlecht, F0 und Sprechgeschwindigkeit). Die Resultate zeigen an, daß weibliche Sprecher als Gesamtgruppe besser verstanden werden als männliche. Weiterhin wurde als Tendenz ermittelt, daß die F0-Variationsbreite positiv mit erhöhten Verständlichkeitswerten korreliert. Allerdings korreliert die Sprachverständlichkeit nicht mit dem F0-Mittelwert oder der Sprechgeschwindigkeit. Anschließend wurden verschiedene detailphonetische akustische Sprechercharakteristika als Korrelate der Sprachverständlichkeit untersucht. Es wurde festgestellt, daß Sprecher mit erweitertem Vokalraum im allgemeinen besser verstanden werden als Sprecher mit reduziertem Vokalraum. Unter genauerer Betrachtung zweier Fälle konsistenter Wahrnehmungsfehler (Tilgung und Silbenzuordnung von

* Corresponding author. E-mail: abradlow@indiana.edu.

[1] Earlier versions of some of this work appeared as part of a contribution to a volume honoring Max Wajskop (Bradlow et al., 1995) and were presented at the 127th meeting of the Acoustical Society of America, at the XIIIth International Congress of Phonetic Sciences, and at the 130th meeting of the Acoustical Society of America.

[2] Audiofiles available. See http://www.elsevier.nl/locate/specom .

PII S0167-6393(96)00063-5

Segmenten) konnte festgestellt werden, daß sich die Wahrnehmungsfehler direkt auf detaillierte Ausprägungen des Timing im Sprachsignal zurückführen lassen. Die Resultate weisen darauf hin, daß sich ein Großteil der Variabilität in der Sprachverständlichkeit auf akustisch-phonetische Charakteristika der jeweiligen Sprecher zurückführen lassen. Eine genaue Erkenntnis dieser Faktoren ist von Nutzen für Verfahren der Sprachsynthese und Spracherkenung, sowie für solche Populationen (z.B. bei Hörbehinderungen oder im Zweitspracherwerb), die besonders sensibel auf Unterschiede in der Verständlichkeit verschiedener Sprecher reagieren.

Résumé

Dans cette étude, on a utilisé une base de données multi-locuteurs contenant des scores d'intelligibilité de 2000 phrases (20 locuteurs, 100 phrases) pour identifier les corrélats de l'intelligibilité qui sont dépendants du locuteur. Nous avons d'abord étudié des caractéristiques "globales" des locuteurs (genre, F0 et vitesse d'élocution). Il est apparu que les locutrices sont globalement plus intelligibles que les locuteurs. Nous avons également observé que l'ampleur du registre de F0 avait tendance à être corrélée positivement avec des scores d'intelligibilité plus élevés. Toutefois, la valeur moyenne de F0 et la vitesse d'élocution ne semblent pas être corrélées avec l'intelligibilité. Nous avons ensuite examiné d'autres corrélats de l'intelligibilité globale, plus fins au niveau des caractéristiques acoustico-phonétiques du locuteur. Nous avons observé que les locuteurs présentant des triangles vocaliques larges étaient généralement plus intelligibles que les locuteurs présentant des triangles vocaliques serrés. En étudiant deux cas d'erreurs d'écoute consistantes (destruction d'un segment et attribution à une syllabe), nous avons trouvé que ces erreurs perceptives pouvaient être dérivées directement des caractéristiques de timing détaillées du signal de parole. Les résultats suggèrent qu'une part substantielle de la variabilité de l'intelligibilité de la parole normale est attribuable aux caractéristiques acoustico-phonétiques spécifiques du locuteur. La connaissance de ces facteurs peut être utile pour améliorer la synthèse de la parole et les stratégies de reconnaissance, et pour des populations spécifiques (comme par exemple, les mal-entendants ou ceux qui apprennent une langue étrangère) qui sont particulièrement sensibles aux écarts d'intelligilibité entre locuteurs.

Keywords: Intelligibility; Talker characteristics; Acoustic-phonetics

1. Introduction

It is well known that even under "ideal" speaking and listening conditions, there is a wide range of individual differences in overall speech intelligibility across normal talkers (e.g. (Black, 1957; Hood and Poole, 1980; Bond and Moore, 1994)). Additionally, recent studies on the role of talker variability in speech perception and spoken word recognition have shown that listeners are sensitive to talker variability to the extent that speech intelligibility scores decrease with increased talker variability in the test materials (e.g. (Mullennix et al., 1989; Pisoni, 1993; Sommers et al., 1994; Nygaard et al., 1995)). Moreover, listeners show evidence of encoding talker-specific voice attributes in memory along with information about the specific test words (Palmeri et al., 1993). Nygaard et al. (1994) also reported that familiarity with a talker's voice leads to an advantage in intelligibility of speech produced by that talker, suggesting a direct link between listener sensitivity to paralinguistic, talker-specific attributes and overall speech intelligibility. Taken together, there is a growing body of research showing that the linguistic content of an utterance and the indexical, paralinguistic information, such as talker- and instance-specific characteristics, are not only simultaneously conveyed by the acoustic signal, but also are not dissociated, or normalized away, during speech perception (Ladefoged and Broadbent, 1957; Laver and Trudgill, 1979).

Similarly, studies of within-talker variability in speech production have shown that talkers systematically alter their speech patterns in response to particular communicative requirements in ways that have substantial effects on the overall intelligibility of an utterance. For example, in a series of studies on speech directed towards the hard of hearing, Picheny et al. (1985, 1986, 1989) and Uchanski et al. (1996) found systematic, acoustic-phonetic differences between "clear" and "conversational" speech within individual talkers. Clear speech had consistently higher intelligibility scores, and was found to be slower and to exhibit fewer phonological reduction

phenomena than conversational speech. Lindblom (1990) and Moon and Lindblom (1994) showed that talkers adapt their speech patterns to both production-oriented and listener-oriented factors as demanded by the specific communicative situation. For example, formant frequencies of vowels embedded in words spoken in "clear speech" exhibited less contextually conditioned undershoot than those embedded in words spoken in "citation form".

Recently, Bond and Moore (1994) investigated whether the acoustic-phonetic characteristics that apparently distinguish "clear" versus "conversational" speaking styles within a talker also distinguish the speech across talkers who differ in overall intelligibility. Indeed, in a comparison of the acoustic-phonetic characteristics of the speech of a relatively high intelligibility talker and two talkers with relatively low intelligibility, Bond and Moore found that "inadvertently" clear speech shared many of the acoustic-phonetic characteristics of intentionally clear speech. Finally, Keating et al. (1994) and Byrd (1994) investigated inter-talker variability in pronunciation of American English from tokens in the TIMIT database of American English dialects (Lamel et al., 1986; Pallett, 1990; Zue et al., 1990). Both of these studies revealed the broad range of pronunciation characteristics in American English, and pointed out how paralinguistic factors, such as the talker's gender, dialect and age, in addition to linguistic factors, such as phonetic context, contribute to the observed pronunciation variability. However, since the TIMIT database does not include perceptual data, neither of these studies could make any inferences regarding the effects of these inter-talker differences on overall speech intelligibility.

The goal of the present study was to extend our understanding of the talker-specific characteristics that lead to variability in speech intelligibility by investigating the acoustic correlates of different talkers' productions in a large database that includes both sentence productions from multiple talkers and intelligibility data from multiple listeners per talker (Karl and Pisoni, 1994). The basic question we asked was: "What acoustic characteristics make some talkers more intelligible than others?" By directly assessing talker-specific correlates of speech intelligibility at the acoustic-phonetic level this investigation aimed to extend our understanding of the relationship between the indexical and linguistic aspects of speech communication: we hoped to identify some of the aspects of talker variability that might, on the one hand, be expected to help identify a particular talker, and on the other hand, have a direct effect on overall speech intelligibility.

We acknowledge that it is misleading to ascribe all of the variability in sentence intelligibility to acoustic-phonetic characteristics of the talker. Such an approach incorrectly disregards any listener-talker-sentence interactions that affect the resultant intelligibility scores. Nevertheless, while keeping in mind the contribution of listener- and sentence-related factors to overall intelligibility, we were interested in investigating what talker-related characteristics, independently of the listener- and sentence-related characteristics, might correlate with overall intelligibility, and therefore might account for some portion of the observed variability in overall intelligibility. We hoped that the results of this investigation combining both acoustic-phonetic measurements with perceptual data might lead to a better understanding of the salient acoustic-phonetic characteristics that listeners respond to during speech perception, and would therefore help to differentiate highly intelligible speech from less intelligible speech.

We adopted an approach that focused on two aspects of talker-specific characteristics. First, we focused on "global" talker characteristics, such as gender, fundamental frequency and rate of speech. These characteristics are "global" because they extend over the entire set of utterances from a given talker, rather than being confined to local aspects of the speech signal that are related to the articulation of individual segments. Second, we focused on specific pronunciation characteristics, such as vowel category realization and segmental timing relations that are fine-grained, acoustic-phonetic indicators of instance-specific variability. Whereas the global characteristics provide information about some of the invariant speech attributes of the individual talkers, the fine-grained acoustic-phonetic details at the local, segmental level, provide information about the instance-specific pronunciation characteristics of particular utterances. We expected that a wide range of these talker-related characteristics would contribute to variability in overall intelligibility, and we hoped that this approach would provide a better understand-

ing of some of the talker- and instance-specific factors that are associated with highly intelligible normal speech.

2. The Indiana multi-talker sentence database

The materials for this study came from the Indiana Multi-Talker Sentence Database (Karl and Pisoni, 1994). This database consists of 100 Harvard sentences (IEEE, 1969) produced by 20 talkers (10 males and 10 females) of General American English [3]. The sentences are all mono-clausal and contain 5 keywords plus any number of additional function words. None of the talkers had any known speech or hearing impairments at the time of recording, and all recordings were live-monitored for gross misarticulations, hesitations, and other disfluencies. (See Table 2 for examples of the sentences.) The sentences were presented to the subjects on a CRT monitor in a sound-attenuated booth (IAC 401A). The stimuli were transduced with a Shure (SM98) microphone, and digitized on-line (16-bit analog-to-digital converter (DSC Model 240) at a 20 kHz sampling rate). The average root mean square amplitude of each of the digital speech files was then equated with a signal processing software package (Luce and Carrell, 1981), and the files were converted to 12-bit resolution for later presentation to listeners in a transcription task using a PDP-11/34 computer.

Along with the audio recordings, this database also includes speech intelligibility data in the form of sentence transcriptions by 10 listeners per talker, for a total of 200 listeners. In collecting these transcriptions, each group of 10 listeners heard the full set of 100 sentences produced by a single talker. The sentence stimuli were low-pass filtered at 10 kHz, and presented binaurally over matched and calibrated TDH-39 headphones using a 12-bit digital-to-analog converter. The listeners heard each sentence in the clear (no noise was added) at a comfortable listening level (75 dB SPL), and then typed what they heard at a computer keyboard. A PDP-11/34 computer was used to control the entire experimental procedure in real-time. The listeners were all native speakers of American English, who were students at Indiana University. They had no speech or hearing impairments at the time of testing.

The sentence transcriptions were scored by a keyword criterion that counted a sentence as correctly transcribed if, and only if, all 5 keywords were correctly transcribed. Any error on a keyword resulted in the sentence being counted as mistranscribed. With this strict scoring method, each sentence for each talker received an intelligibility score out of a possible 10. Each talker's overall intelligibility score was then calculated as the average score across all 100 sentences.

As shown in Table 1, the overall sentence intelligibility scores ranged from 81.1% to 93.4% correct transcription, with a mean and standard deviation of 87.8% and 3.1%, respectively. Thus, the materials in this large multi-talker sentence database showed considerable variation and covered a range of talker intelligibility that could be used as the basis for an investigation of the effects of global and fine-grained acoustic-phonetic talker characteristics on overall speech intelligibility.

It is important to note here that intelligibility scores must be interpreted in a relative sense. For example, Hirsh et al. (1954) observed that authors on this subject almost always caution readers "... to regard such scores as specific to a given crew of talkers and a given crew of listeners". In the present study, we were specifically interested in exploring the individual characteristics of our "crew of talkers", however, our database was constructed in such a way that it did not provide the means of systematically investigating the contribution of the "crew of talkers" independently of the "crew of listeners". This is because for each talker, a different group of 10 listeners, drawn from the same population, transcribed the recordings of the full set of 100 sentences. Therefore, the intelligibility scores for the 20 talkers shown in Table 1, as well as the talker-related correlates of intelligibility that we discuss below, should, strictly speaking, be regarded as reflecting characteristics of the particular talker-listener situation, rather than of the talker independently of

[3] Copies of the Indiana Multi-Talker Sentence Database can be obtained in CD-ROM form for a nominal cost for media and postage. Please write the authors at Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405, USA, or e-mail, abradlow@indiana.edu.

Table 1
Intelligibility scores across all 100 sentences, F0 mean, F0 minimum, F0 maximum, F0 range, and mean sentence duration for each individual talker. Asterisks mark the two talkers (F7 and M2) whose vowel space measurements are shown in Fig. 2, and whose speech samples can be heard from the Elsevier web site (http://www.elsevier.nl/locate/specom)

Talker	Intelligibility (100 sentences)	F0 mean (Hz)	F0 min. (Hz)	F0 max. (Hz)	F0 range (Hz)	Mean sentence duration (sec.)
F1	88.4	208.4	86.64	305.58	218.9	2.8
F2	91.5	168.4	107.91	241.9	134.0	1.9
F3	89.6	178.5	97.08	247.3	150.2	1.9
F4	87.3	210.7	129.14	293.66	164.5	1.8
F5	90.1	220.7	96.01	322.85	226.8	1.9
F6	87.8	203.3	80.08	311.12	231.0	2.2
F7*	93.4	162.8	90.4	230.1	139.7	2.2
F8	87.4	206.9	152.25	276.71	124.5	1.9
F9	88.2	206.9	123.26	281.58	158.3	2.2
F10	91.0	237.3	119.97	325.6	205.6	2.0
M1	88.8	119.6	63.34	177.79	114.4	2.3
M2*	81.1	141.8	79.08	210.89	131.8	2.1
M3	84.8	130.3	81.31	177.31	96.0	2.8
M4	81.7	102.5	68.3	144.87	76.6	1.9
M5	88.9	110.2	64.67	164.82	100.1	2.1
M6	89.0	140.3	81.31	221.39	140.1	2.0
M7	89.8	100.0	66.99	140.36	73.4	2.3
M8	87.2	118.7	66.23	169.64	103.4	2.0
M9	87.3	118.7	69.13	184.95	115.8	2.1
M10	83.5	104.0	70.85	150.4	79.5	1.9

the listener. Nevertheless, our operating assumption was that if we could trace some of the observed variability in intelligibility to acoustic-phonetic characteristics of the speech signal, then we would indeed be tapping into some of the talker-dependent correlates of intelligibility that might reasonably be expected to affect overall intelligibility for a wide range of listeners.

3. Global talker characteristics

3.1. Gender

We began by investigating global talker characteristics that could provide an indication of the relationship between source-related acoustic characteristics and overall speech intelligibility scores. Although all of the talkers in our database were judged to have normal voice qualities, we investigated whether some voice qualities would be associated with higher speech intelligibility scores than others. In particular, we wondered whether the talker's gender would be a correlate of variability in intelligibility.

Male and female glottal characteristics differ considerably (Klatt and Klatt, 1990; Hanson, 1995), and listeners are generally able to distinguish male from female voices quite easily (Nygaard et al., 1994; Tielen, 1992). Furthermore, Byrd (1994) found that male speech in the TIMIT database of American English was characterized by a greater prevalence of phonological reduction phenomena, such as vowel centralization, alveolar flapping, and reduced frequency of stop releases, relative to female speech in this database. Thus, there is some evidence that gender is a salient characteristic that we might expect to affect overall intelligibility. Specifically, we hypothesized that more "reduced" speech would lead to lower overall intelligibility, and therefore that the group of female talkers in our database might have a higher mean overall intelligibility score than the group of male talkers. Indeed, we found that the group of 10 female talkers did have a significantly higher overall intelligibility score than the group of 10 male talkers (89.5% versus 86.2% correct transcription with standard deviations of 2.0% and 3.2%, respectively; $t(18) = 2.72$, $p = 0.01$ by a 2-tail unpaired t-test). Furthermore, the four talkers with the

highest overall intelligibility scores were female and the four talkers with the lowest overall intelligibility scores were male (see Table 1). This result raised the question of what specific acoustic-phonetic characteristics lead to this gender-based intelligibility difference. Byrd's analyses suggest that this intelligibility difference might be due to an increased prevalence of specific reduction phenomena for male speech relative to female speech, rather than due to the source-related (voice quality) differences between males and females (Byrd, 1994). However, before turning to a discussion of fine-grained pronunciation differences, we examined several other global talker characteristics that might provide information about the relationship between talker-specific factors and overall speech intelligibility.

3.2. Fundamental frequency

Fundamental frequency is a global talker characteristic that typically differs markedly across male and female talkers. However, it is not clear that it is an acoustic attribute that directly affects overall speech intelligibility. Bond and Moore (1994) found no reliable difference in mean fundamental frequency between their higher and lower intelligibility talkers. Similarly, Picheny et al. (1986) found that for all three talkers in their study, clear speech was characterized by a somewhat wider range in fundamental frequency with a slight bias towards higher fundamental frequencies than conversational speech, however, these differences were not dramatic. In the present study, we investigated both fundamental frequency mean and range as possible correlates of overall intelligibility; however, based on these previous studies, we had no strong predictions regarding the relationship between fundamental frequency characteristics and intelligibility.

All fundamental frequency measurements were made using the Entropics WAVES + software (version 5.0) on a SUN workstation. For each sentence produced by each talker, the mean, minimum and maximum fundamental frequency was extracted from the voiced portions of the digital speech file using the pitch extraction program included in the Entropics WAVES + software package. Each talker's overall mean, minimum and maximum fundamental frequency was then calculated across all 100 sentences. These values are given in Table 1.

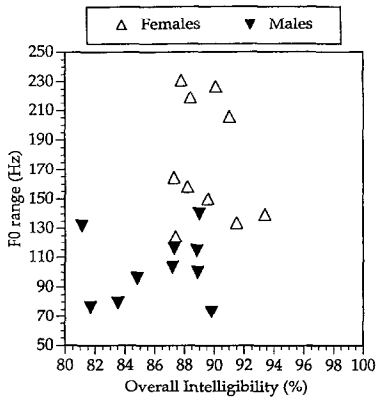

Fig. 1. Scatter plot of overall intelligibility as a function of F0 range. Male talkers are represented by closed triangles; female talkers are represented by open triangles.

Because the female talkers taken as a group had a higher mean intelligibility score, and as expected the female talkers had a higher mean fundamental frequency, across all talkers we found a slight tendency for a higher mean fundamental frequency to correlate with a higher mean intelligibility score (Spearman [4] $\rho = +0.341$, $p = 0.14$). However, when we looked at the males and females separately, we found no such correlation between mean fundamental frequency and intelligibility. Thus, in our database, overall intelligibility is not correlated with mean fundamental frequency independently of the gender-based difference in overall intelligibility. With respect to fundamental frequency range, across all 20 talkers we found a tendency for a wider range in fundamental frequency to correlate with a higher overall intelligibility score (Spearman $\rho = +0.384$, $p = 0.095$). Fig. 1 shows a scatter plot of fundamental frequency range (in Hertz) against overall intelligibility for the male talkers (closed triangles) and female talkers (open triangles).

[4] Throughout this paper, we report Spearman rank order correlation coefficients rather than Pearson product-moment correlation coefficients. The use of this non-parametric correlation coefficient is appropriate in cases where at least one of the scales is ordinal (Runyon and Haber, 1991). Since our speech intelligibility scores constitute a relative scale, along which the 20 talkers are ranked, Spearman rank order correlations were deemed appropriate.

We also found a significantly greater fundamental frequency range for the group of female talkers than for the group of male talkers ($t(18) = 4.87$, $p < 0.001$ by a 2-tailed unpaired t-test.) The mean and standard deviation of the fundamental frequency range were 175 Hz and 41 Hz for the female group, and 103 Hz and 23 Hz for the male group, respectively. Since this finding is correlational, we cannot be certain whether the wider fundamental frequency range leads to higher intelligibility or whether both the higher intelligibility and wider fundamental frequency range are simply consequences of some other voice quality attribute of our female talkers. One piece of evidence that bears on this issue comes from a recent study by Tielen (1992) who found that, although female speakers of Dutch typically had higher mean fundamental frequencies than their male counterparts, they did not have significantly wider fundamental frequency ranges. For the purposes of the present study, this finding indicates that a wider fundamental frequency range is not a necessary consequence of a higher fundamental frequency mean. It is therefore possible that the wider female fundamental frequency range is one of the female speech characteristics that contributes to the generally higher intelligibility of female speech relative to male speech in our database.

3.3. Speaking rate

The final global talker characteristic that we investigated was overall speaking rate. Although speaking rate is not a source-related, voice-quality characteristic, it is one of the most salient global talker-specific characteristics, and one that is known to distinguish "clear" versus "conversational" speech within individuals (Picheny et al., 1989; Krause and Braida, 1995; Uchanski et al., 1996). Additionally, many phonological reduction phenomena are directly related to changes in speaking rate. In Byrd's analyses of the TIMIT database, which included sentences from 630 talkers, she found that across all dialects, the males had significantly faster speaking rates than the females on the two calibration sentences that were read by all talkers. However, Byrd's study also found an interaction of gender and dialect region such that the slowest speaking region for the male speakers (the South Midland) was only the fourth slowest for the female speakers. Bond and Moore (1994) found no word duration differences in their analyses of two talkers that differed in overall intelligibility when the words were embedded in sentences, although for isolated words the less intelligible talker had shorter durations than the more intelligible talker. Furthermore, in a recent study of the effects of speaking rate on the intelligibility of clear and conversational speaking modes, Krause and Braida (1995) reported that trained talkers were able to achieve an intelligibility advantage for the clear speech mode even at faster speaking rates. In other words, it is possible to produce fast clear speech. Thus, although there is some evidence that overall speaking rate varies with paralinguistic (indexical) characteristics such as speaker's gender and dialect, and that speaking rate can be associated with a "reduced", conversational speaking style, a direct link between speaking rate and intelligibility remains unclear.

In our database, we measured overall speaking rate for each of the 20 talkers, as the mean sentence duration across all 100 sentences. All duration measurements were made using the Entropics WAVES + software on a SUN workstation. The questions we asked here were: (1) Does overall speaking rate correlate with overall speech intelligibility across all 20 talkers? and (2) Can the gender-based intelligibility difference be traced to a gender-based difference in overall speaking rate? As shown in Table 1, we observed considerable variability across all 20 talkers in mean sentence duration (mean sentence duration = 2.115 seconds, with a standard deviation of 0.276 seconds). However, we failed to find a clear relationship between mean sentence duration and overall speech intelligibility scores: there was no correlation between speaking rate and speech intelligibility across all 20 talkers, and there was no significant difference in the means between the male and female speaking rates.

Thus, in our multi-talker sentence database, overall speaking rate as measured by mean sentence duration did not appear to be a talker-related correlate of variability in speech intelligibility. This result is consistent with the recent finding of Krause and Braida (1995) that fast speech can also be "clear" speech, and with Bond and Moore (1994) who found no difference in duration for words in sentences

spoken by a high and a low intelligibility talker. Furthermore, even though the present data do not show a difference between male and female speaking rates as reported by Byrd (1994), it is likely that these data reflect the interaction between speaker gender and dialect region that she found in the TIMIT database: most of our speakers and listeners were from the South Midland region, which Byrd found to have the slowest speaking rate for males but an average speaking rate for females. Nevertheless, it remains possible that a measure of speaking rate that took into account the number and duration of any pauses that the talker may have inserted into the sentence, rather than simply averaging the pauses into the overall sentence durations, would correlate better with overall intelligibility. This possibility is supported by the finding of Picheny et al. (1986, 1989) and Uchanski et al. (1996) who reported that, within individual talkers, clear speech contains more numerous and longer pauses than conversational speech.

4. Fine-grained acoustic-phonetic talker characteristics

4.1. Vowel space characteristics

We began our investigation of fine-grained acoustic-phonetic talker characteristics with an examination of vowel spaces. Vowel centralization is a typical feature of casual, or reduced speech (Picheny et al., 1986; Lindblom, 1990; Moon and Lindblom, 1994; Byrd, 1994). Additionally, vowel space expansion has been shown to correlate with speech intelligibility. For example, Bond and Moore (1994) found more peripheral vowel category locations in an F1 by F2 space for a higher-intelligibility talker relative to a lower-intelligibility talker. In a study of vowel production by deaf adolescents, Monsen (1976) found a significant positive correlation between range in F2 and intelligibility. Both of these studies lead us to hypothesize that in our multi-talker sentence database we would find a positive correlation between overall intelligibility and measures of vowel space expansion. Specifically, we predicted that relatively expanded vowel spaces would be associated with enhanced speech intelligibility scores.

Table 2
Subset of 18 sentences containing the words with the target vowels from which the vowel space measurements were taken, with the IPA phonemic transcription for the target word. All 5 keywords are italicized, with the word with the target vowel in boldface. Asterisks mark the three sentences whose productions by Talker F7 and M2 can be heard from the Elsevier web site (http://www.elsevier.nl/locate/specom)

/i/:	
1*. *It's* ***easy*** to *tell* the *depth* of a *well.*	/izi/
2. The *fruit* ***peel*** was *cut* in *thick slices.*	/pil/
3. *Adding fast* ***leads*** to *wrong sums.*	/lidz/
4. *This* is a *grand* ***season*** for *hikes* on the *road.*	/sizIn/
5. The *walled town* was ***seized*** without a *fight.*	/sizd/
6. The ***meal*** was *cooked before* the *bell rang.*	/mil/
/a/:	
7*. A ***pot*** of *tea helps* to *pass* the *evening.*	/pat/
8. A ***rod*** is *used* to *catch pink salmon.*	/rad/
9. The *wide road shimmered* in the ***hot*** sun.	/hat/
10. The *show* was a ***flop*** from the *very start.*	/flap/
11. The *hogs* were *fed* ***chopped*** *corn* and *garbage.*	/tʃapt/
12. A *large size* in ***stockings*** is *hard* to *sell.*	/stakɪŋz/
/o/:	
13*. The *horn* of the *car* ***woke*** the *sleeping cop.*	/wok/
14. *Bail* the ***boat*** to *stop* it *from sinking.*	/bot/
15. *Mend* the ***coat*** *before* you *go out.*	/kot/
16. *Hoist* the ***load*** to *your left shoulder.*	/lod/
17. The *dune* ***rose*** *from* the *edge* of the *water.*	/roz/
18. The *young girl gave* ***no*** *clear response.*	/no/

In order to measure each talker's vowel space, we selected six occurrences of the three peripheral vowels, /i,a,o/, from the sentence materials in the database. (The point vowel /u/ was avoided due to excessive allophonic variation for this vowel in General American English.) All of the words containing the target vowels were content words, and none was the final keyword in the sentence. Table 2 lists the subset of 18 sentences containing the words with the target vowels from which the vowel space measurements were taken.

The first and second formants were measured from each of the 18 target vowels as produced by each of the 20 talkers. All formant measurements were made using the Entropics WAVES + software package on a SUN workstation. Both LPC spectra (calculated from a 25 ms Hanning window) and spectrograms were used to determine the location of the first two formant frequencies at the vowel steady-state. These F1 and F2 measurements were then converted to the perceptually motivated mel

Table 3
Intelligibility scores across the 18 sentences used for the vowel space measurements, vowel space area, vowel space dispersion, F1 range, F2 range, within category clustering, vowel space dispersion/within-category clustering, F2–F1 distance for /i/ and /a/ for each individual talker. Asterisks mark the two talkers (F7 and M2) whose vowel space measurements are shown in Fig. 2, and whose speech samples can be heard from the Elsevier web site (http://www.elsevier.nl/locate/specom)

Talker	Intelligibility (18 sentences)	Vowel space area ($mels^2$)	Vowel space dispersion (mels)	F1 range (mels)	F2 range (mels)	Category clustering (mels) clustering	Vowel space dispersion/ category	F2–F1 /i/ (mels)	F2–F1 /a/ (mels)
F1	93.3	82747.76	349.23	649.72	802.95	66.899	5.220	1426.44	311.19
F2	92.8	40844.95	301.91	569.19	746.18	75.077	4.021	1396.64	339.11
F3	91.1	81688.80	327.33	518.35	1168.76	90.939	3.599	1314.26	412.39
F4	85.0	49686.44	268.21	545.66	932.08	110.827	2.420	1219.59	422.63
F5	91.1	85160.25	311.06	604.91	1172.23	107.85	2.884	1357.83	496.94
F6	91.7	69203.79	321.54	542.88	852.91	67.482	4.765	1370.43	333.86
F7*	92.8	98726.79	346.85	607.76	904.66	62.253	5.572	1394.58	307.04
F8	92.8	55770.36	291.36	572.61	757.56	77.757	3.747	1351.63	405.06
F9	88.3	28993.41	251.30	564.30	651.23	73.608	3.414	1243.95	339.65
F10	90.6	61950.01	252.13	472.95	672.00	66.179	3.810	1253.27	450.47
M1	91.1	61092.87	285.57	470.40	844.43	61.688	4.629	1230.42	355.51
M2*	78.3	41005.79	272.45	435.03	737.49	65.893	4.135	1282.32	413.46
M3	82.2	114352.73	360.09	498.13	1053.08	94.095	3.827	1238.78	393.10
M4	81.5	73394.01	278.47	456.37	745.809	65.757	4.235	1211.41	529.38
M5	85.0	13531.00	250.13	476.60	636.17	55.574	4.501	1268.43	375.96
M6	86.7	72398.30	280.60	475.77	663.73	47.559	5.900	1250.89	443.88
M7	88.3	35205.43	273.67	408.63	811.51	45.863	5.967	1204.81	482.01
M8	90.6	49982.49	262.61	430.11	751.19	99.5	2.639	1080.55	382.81
M9	86.7	63413.41	263.44	453.04	756.13	66.161	3.982	1177.70	387.12
M10	85.0	79670.34	309.08	575.71	878.97	72.69	4.252	1343.15	399.74

scale (Fant, 1973). (The exact equation for converting frequencies from Hertz to mels is $M = (1000/\log 2)\log((F/1000)+1)$, where M and F are the frequencies in mels and Hertz, respectively.) Each talker's vowel space was then represented by the locations of the 18 individual vowel tokens in an F1 by F2 space. In all of the following analyses of the relations between vowel space characteristics and speech intelligibility, we used each talker's average intelligibility score across the 18 sentences, given in Table 2, that formed the subset of sentences with the words that contained the target vowels (see Table 3). Across all 20 talkers, the overall intelligibility scores for the total set of 100 sentences and for the subset of 18 sentence were significantly correlated (Spearman $\rho = +0.629$, $p = 0.006$), thus this subset of 18 sentences was a good indicator of the talkers' overall intelligibility scores.

The first measure that we used to assess the relationship between vowel space and overall speech intelligibility was the Euclidian area covered by the triangle defined by the mean of each vowel category. Here we hypothesized that the greater the triangular area, the higher the overall intelligibility. Fig. 2(a) shows the vowel triangles for the highest intelligibility talker (Talker F7) and the lowest intelligibility talker (Talker M2). Sample sentences that provide an indication of these two talkers' vowel spaces can be heard in Signals [5] A–F (three sentences for each talker). It is clear from Fig. 2(a) that the vowel triangle for Talker F7 covers a greater area within this space than the vowel triangle for Talker M2. However, across all 20 talkers we failed to find a positive correlation between triangular vowel space

[5] The texts corresponding to Signals A–J are given in Appendix A.

area and speech intelligibility scores (see Table 3 for each individual talker's vowel space area). One problem with triangular vowel space area as a measure of vowel category differentiation is that the points used to calculate this measure are the category averages, and these may not be representative of the individual

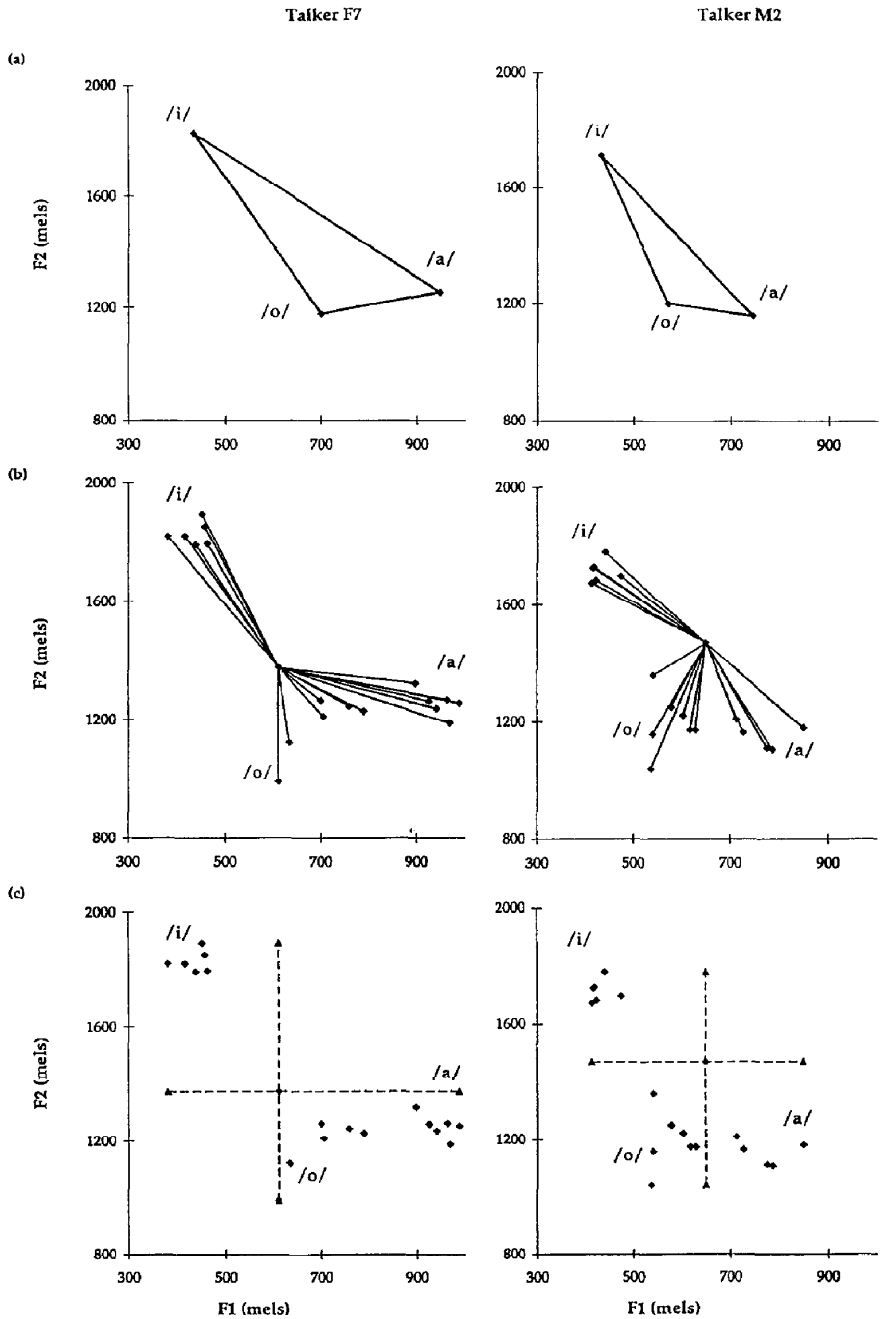

Fig. 2. Vowel space characteristics for a high-intelligibility talker (Talker F7) and a low-intelligibility talker (Talker M2): (a) vowel space area, (b) vowel space dispersion, (c) range in F1 and F2.

vowel tokens actually produced by the talker. For this reason, we devised a different measure of vowel space expansion that took into account the specific location of each individual vowel token, and then reanalyzed the data.

Fig. 2(b) shows each vowel token's distance from a central point in the talker's vowel space for the highest intelligibility talker (Talker F7) and the lowest intelligibility talker (Talker M2). A measure of each talker's "vowel space dispersion" was calculated as the mean of these distances for each talker. This measure thus provided an indication of the overall expansion, or compactness, of the set of individual vowel tokens from each talker (see Table 3 for each individual talker's vowel space dispersion measure). The measures of vowel space area and vowel space dispersion were highly correlated (Spearman $\rho = +0.782$, $p < 0.001$), however, the correlation was not perfect indicating that each measure captures a slightly different aspect of the talkers' vowel production characteristics. With respect to the correlation between vowel space dispersion and intelligibility, we found a moderate, positive rank order correlation (Spearman $\rho = +0.431$, $p = 0.060$) across all 20 talkers, and this correlation increased when only the 10 highest intelligibility talkers were included in the analysis (Spearman $\rho = +0.698$, $p = 0.036$). Thus, using a measure of vowel space dispersion, the data showed that higher overall speech intelligibility is associated with a more expanded vowel space, particularly for the talkers in the top half of the distribution of intelligibility scores.

Based on the finding that overall vowel space dispersion and speech intelligibility were correlated, we then investigated which of the two dimensions, F1 or F2, in the vowel space representations was more responsible for this correlation. In his study of the vowel productions of deaf adolescents, Monsen (1976) found a stronger positive correlation between range in F2 and intelligibility ($r = +0.74$) than he did for range in F1 and intelligibility ($r = +0.45$). As Monsen notes, these correlations do not suggest that range in F2 is more important for normal speech intelligibility than range in F1, rather these correlations arise from the fact that the vowels of these deaf subjects occupy a more normal range in F1 than in F2. For the purposes of our investigation of variability in normal speech, Monsen's finding simply indicated the usefulness of investigating range in F1 and F2 as separate dimensions that might correlate with overall intelligibility.

Accordingly, we measured each talker's range in F1 and F2 as the difference between the maximum and minimum values on each of these dimensions. Fig. 2(c) shows the F1 and F2 range measurements for the highest intelligibility talker (Talker F7) and the lowest intelligibility talker (Talker M2). (See Table 3 for each individual talker's range in F1 and F2.) Across all 20 talkers, we found a significant positive rank order correlation between range in F1 and intelligibility (Spearman $\rho = +0.531$, $p = 0.020$), but we failed to find a significant rank order correlation between range in F2 and intelligibility (Spearman $\rho = +0.239$, $p = 0.300$). This correlation of F1 range and intelligibility was strengthened when only the top 10 talkers were included in the analysis (Spearman $\rho = +0.817$, $p = 0.014$). Thus, it appears that the area covered in F1 was a better correlate of overall intelligibility than the area covered in F2. This finding is not surprising in view of the fact that the English vowel system has several vowel height distinctions (of which F1 frequency is an important acoustic correlate), whereas there are many fewer distinctions along the front–back dimension (of which F2 frequency is the primary acoustic correlate). It may be that in order for the numerous English vowels to be well distinguished, a wide range in F1 (vowel height) is advantageous, whereas less precision can be more easily tolerated in the F2 (front–back) dimension.

The vowel space measures that we have reported so far have established relations between relative vowel space expansion and overall speech intelligibility, particularly for talkers in the top half of the intelligibility score distribution. An additional measure of vowel articulation that might be expected to correlate with intelligibility is the relative compactness of individual vowel categories. We might expect that the more tightly clustered categories enhance intelligibility since they are less likely to lead to inter-category confusion. As a measure of tightness of within-category clustering, we first calculated the mean of the distances of each individual token from the category mean, as we did for our measure of overall vowel space dispersion. Then a single measure for each talker was calculated as the mean

within-category dispersion across all three vowel categories (see Table 3 for these values for each talker). However, analysis of the results showed that across all 20 talkers, as well as for only the 10 highest intelligibility talkers, there was no correlation between within-category dispersion and intelligibility. Thus, tightness of within-category clustering per se was not a good correlate of overall intelligibility.

We then explored the possibility that a combined measure of within- and between-category dispersion might correlate with intelligibility better than each measure independently. We hypothesized that talkers can compensate for a less dispersed overall vowel space by having more tightly clustered individual vowel categories. In order to test this hypothesis we calculated a "dispersion index" from each talker's overall vowel space dispersion divided by the mean within-category clustering (see Table 3). We expected that a greater dispersion index would indicate better differentiated vowel categories relative to the overall vowel space area, and would therefore correlate positively with overall intelligibility. Across all 20 talkers, the dispersion index did not correlate with intelligibility; but, for the 10 highest intelligibility talkers there was a significant positive correlation (Spearman $\rho = +0.654$, $p = 0.049$). However, this correlation is comparable to the correlation between overall vowel space dispersion and intelligibility independently of within-category clustering (Spearman $\rho = +0.698$, $p = 0.036$), suggesting that overall vowel space expansion on its own, rather than relative to within-category compactness, is associated with increased speech intelligibility.

The final measure of vowel space that we examined as a possible correlate of speech intelligibility was the acoustic-phonetic implementation of the point vowels /i/ and /a/. Each of these two vowel categories defines an extreme point in the American English general vowel space. In the acoustic domain, they each display extreme F2–F1 distances: /i/ is characterized by a wide separation between the first two formant frequencies, whereas /a/ is characterized by very close F1 and F2 frequencies. Thus, the F2–F1 distance for these point vowels provided an indication of the extreme locations in the F1 by F2 space for these vowels (Gerstman, 1968). Accordingly, we hypothesized that the F2–F1 distance for /i/ would be positively correlated with overall intelligibility, and that the F2–F1 distance for /a/ would be negatively correlated with overall intelligibility. Indeed, across all 20 talkers, we found a positive rank order correlation between F2–F1 distance for /i/ and overall intelligibility (Spearman $\rho = +0.601$, $p = 0.009$), and a negative rank order correlation between F2–F1 distance for /a/ and overall intelligibility (Spearman $\rho = -0.509$, $p = 0.027$). (See Table 3 for these values for each talker.) When only the 10 highest intelligibility talkers were included in the analysis, these correlations were strengthened further (Spearman $\rho = +0.866$, $p = 0.009$ and Spearman $\rho = -0.673$, $p = 0.043$, for /i/ and /a/, respectively). Thus, relatively high overall speech intelligibility is associated with more extreme vowels as measured by the precision of individual vowel category realization, as well as by overall vowel space expansion for a given talker.

In summary, the general pattern that emerged from these measures of the acoustic-phonetic vowel characteristics as correlates of overall intelligibility was that talkers with more reduced vowel spaces tended to have lower overall speech intelligibility scores. The measures of vowel space reduction that were shown to correlate with overall speech intelligibility were overall vowel space dispersion, particularly range covered in the F1 dimension, and the extreme locations in the F1 by F2 space of the point vowels /i/ and /a/ as measured by F2–F1 distance. The analyses also showed that the correlations between vowel space reduction and overall intelligibility were stronger for talkers in the top half of the distribution of intelligibility scores, suggesting a greater degree of variability for talkers with lower intelligibility scores that is not accounted for by these measures of a talker's vowel space.

4.2. Acoustic-phonetic correlates of consistent listener errors

Another strategy we used for investigating the correlation between fine-grained acoustic-phonetic characteristics of a talker's speech and overall intelligibility involved analyses of the specific portions of sentences that showed consistent listener transcription errors. With this approach we hoped to identify specific pronunciation patterns that resulted in the observed listener errors. These analyses differed from

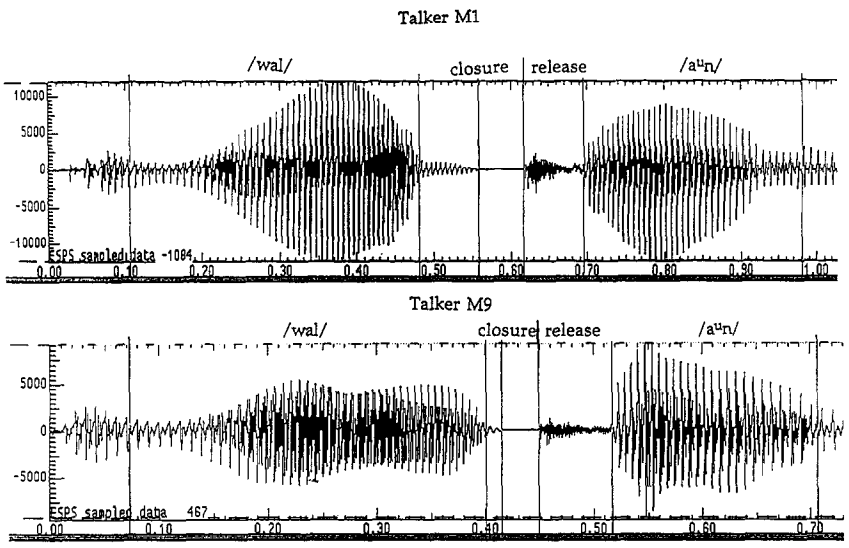

Fig. 3. Waveforms of the sentence portion, "walled town", as produced by Talker M1, who had a relatively long duration of voicing during the stop closure, and Talker M9, who had a very short duration of voicing during the stop closure.

the methods used in the analysis of vowel spaces because here we focused on specific cases where there were known listener errors, rather than on more general statistical indicators of overall phonetic reduction. In particular, in our database we found two specific cases of consistent listener error that revealed the importance of highly precise inter-segmental timing for speech intelligibility (see also (Neel, 1995)).

4.2.1. Segment deletion

The first case of consistent listener error occurred in the sentence "The walled town was seized without a fight". The overall intelligibility of this sentence across all 20 talkers was 60% correct, with 94% of the listener transcription errors occurring for the phrase "walled town". Of the listener errors on this portion of the sentence, 82% involved omitting the word final /d/ in "walled". (None of the remaining 18% of the errors involved omitting the word initial /t/ in "town".) In order to determine what specific talker-related acoustic characteristics might lead to this common listener error, we measured the durations of various portions of the acoustic waveform from this phrase, and then correlated these measurements with the rate of /d/ detection for each talker.

We began by measuring the total vowel-to-vowel duration, that is, the portion of the waveform that corresponds to the talker's /dt/ articulation between the /al/ of "wall" and the /aʊ/ of "town". This portion of the acoustic signal was measured from the point at which there was a marked decrease in amplitude and change in waveform shape as the preceding vowel-sonorant sequence (the /al/ from "wall") ended, until the onset of periodicity for the following vowel (the /aʊ/ from "town"). In almost all cases, this portion consisted of a single closure portion and a single release portion: most talkers (18/20) did not release the /d/ and then form a second closure for the /t/. Fig. 3 shows waveforms of this portion of the sentence for two talkers, with vertical cursors demarcating the salient acoustic boundaries. These sentences can be heard in Signal G (Talker M1) and Signal H (Talker M9).

Across the group of 20 talkers, we found a significant positive rank order correlation between the vowel-to-vowel duration and rate of /d/ detection (Spearman $\rho = +0.713$, $p = 0.002$) [6]. Based on this

[6] Note that the correlations reported here differ slightly from those reported in (Bradlow et al., 1995). This minor difference is due to the addition of one more listener's data into the present analysis: the earlier report was based on only 199 (instead of 200) listeners' data.

finding, we then looked at the rate of /d/ detection in relation to the separate durations of the closure portion and of the release portion, which together comprised the vowel-to-vowel portion. Here we found a significant positive correlation with the closure duration (Spearman $\rho = +0.641$, $p = 0.005$), but no correlation with the release duration. The closure portion generally consisted of a period with very low amplitude, low frequency vibration, followed by a silent portion. Accordingly, we then examined the correlation between rate of /d/ detection and the separate durations of each of these portions of the total closure duration. A highly significant positive correlation was found between rate of /d/ detection and the duration of voicing during the closure (Spearman $\rho = +0.755$, $p < 0.001$), whereas no correlation was found between the duration of the silent portion of the closure and rate of /d/ detection. This correlation suggests that the duration of voicing during closure, in an absolute sense, is a reliable acoustic cue to the presence of a voiced consonant in this phonetic environment. However, an extremely strong (and highly significant) rank order correlation was found between the rate of /d/ detection and the duration of the voicing during the closure relative to the duration of the preceding vowel–sonorant sequence, /wal/ (Spearman $\rho = +0.810$, $p = 0.0004$). In other words, listeners appeared to rely heavily on relative timing between the duration of voicing during the closure and the overall rate of speech, as determined by the duration of the preceding syllable portion, in detecting the presence or absence of a segment. This finding is consistent with studies on rate-dependent processing in phonetic perception that have shown that listeners adjust to overall rate of speech in the identification of phonetic segments (e.g. (Miller, 1981)), and that relative timing between segments can play a crucial role in segment identification (Port, 1981; Port and Dalby, 1982; Parker et al., 1986; Kluender et al., 1988).

Fig. 3 contrasts two talkers with varying amounts of this low frequency voicing during the closure relative to the preceding /wal/ portion of the waveform. Talker M1 had a considerably longer relative duration of voicing during the closure than talker M9, and consequently all of the listeners for Talker M1 detected the presence of the /d/, whereas only 1 of the 10 listeners for Talker M9 detected the /d/. In other words, talkers who did not produce sufficient voicing during the closure as determined by the overall rate of speech (e.g. Talker M9 in Fig. 3) were likely to be mis-heard, even though the syntactic structure of the phrase should have lead listeners to expect "*walled* town" rather than "*wall* town". For the purposes of the present investigation, this particular case of consistent listener-error revealed just how some of the observed variability in speech intelligibility scores for these sentences can be traced directly to specific pronunciation characteristics of the talker. Furthermore, this case indicates the importance of articulatory precision in the realization of gestural timing relations for speech intelligibility.

4.2.2. *Syllable affiliation*

The second case of a consistent listener error occurred in the sentence "The play seems dull and quite stupid". The overall intelligibility of this sentence across all 20 talkers was 75% correct, with 70% of the listener transcription errors occurring for the phrase "play seems". Of the perceptual errors on this portion of the sentence, 95% involved mis-syllabification of the word initial /s/, resulting in "place seems". In this case, we measured the duration of the /s/ (marked by the high frequency, high amplitude turbulent waveform) and of the preceding and following syllables (/plej/ and /simz/, respectively). Fig. 4 shows waveforms of this portion of the sentence for Talker F6 and Talker F1 with vertical cursors marking these three segments. These sentences can be heard in Signal I (Talker F6) and Signal J (Talker F1).

We then examined the correlation of these durations and the rate of correct transcription of "play". We expected to find a correlation between /s/ duration relative to the durations of surrounding syllables and rate of correct transcription. Indeed, results showed a significant negative correlation between rate of "play seems" transcription and /s/ duration as a proportion of the preceding syllable, /plej/, duration (Spearman $\rho = -0.631$, $p = 0.006$). We also found a tendency for the correct transcription rate to correlate with /s/ duration as a proportion of the following syllable, /imz/, duration (Spearman $\rho = -0.432$, $p = 0.060$). In other words, the shorter the /s/ relative to the surround-

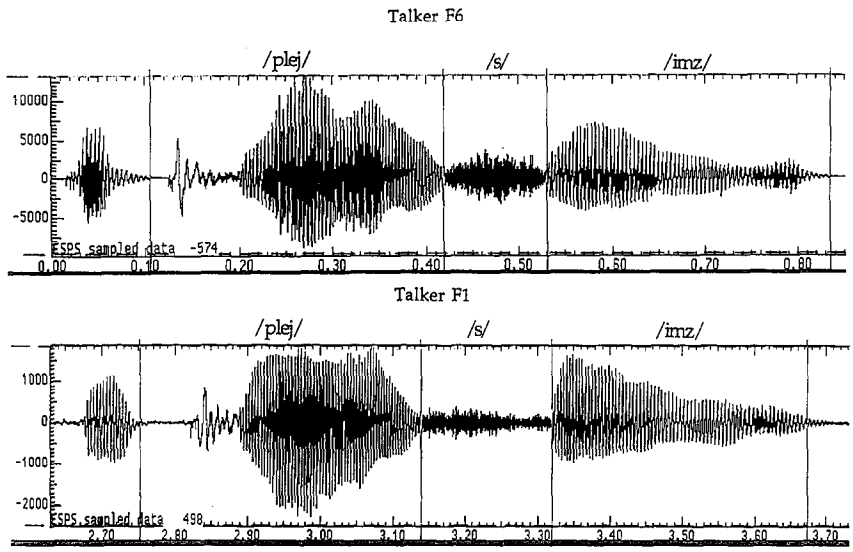

Fig. 4. Waveforms of the sentence portion, "play seems", as produced by Talker F6, who had a short /s/ duration relative to the durations of the /plej/ and /imz/ portions, and Talker F1, who had a long /s/ duration relative to the durations of the /plej/ and /imz/ portions.

ing syllables, the more likely it was to be correctly syllabified by the listener as onset of the following syllable, rather than as both coda of the preceding word and the onset of the following word. In Fig. 4, this may be seen by the shorter relative durations of the /s/ for Talker F6, whose /s/ was correctly syllabified by all 10 listeners, as opposed to the relatively longer /s/ for Talker F1, whose /s/ was correctly syllabified by only 3 of the 10 listeners. Thus, in this case, as in the case of segment deletion discussed above, the listeners drew on global information about the speaking rate of the talker in perceiving the placement of the word boundary. The talker's precision in inter-segmental timing had a direct effect on the listener's interpretation of the speech signal.

Furthermore, in this case, there was a gender-related factor in the timing relationship between the medial /s/ and the surrounding syllables. In general, the duration of the /s/ relative to the preceding and following syllables was shorter for the female talkers than for the male talkers. Consequently, the female talkers' renditions of this phrase were more often correctly transcribed: 7 of the 10 female talkers had no errors of this type, whereas 6 of the 10 males had this error for at least 30% of the listeners. Thus, in this case, the female talkers as a group were apparently more precise with respect to controlling this timing relationship than the group of male talkers. Although this case is not a matter of phonological reduction (in fact, the correct form is shorter in duration), this example does demonstrate that the gender-based difference in overall speech intelligibility that we observed in our database may be due to the use of more precise articulations by our female talkers. Moreover, both this case of syllable affiliation and the previous case of segment deletion indicate why global talker-related characteristics, such as overall speech rate, may not be good candidates for the primary determiners of talker intelligibility: apparently, finer acoustic-phonetic details of speech timing and the precision of specific articulatory events "propagate up" to higher levels of processing during speech perception to modulate and control overall speech intelligibility in sentences.

5. General summary and discussion

The overall goal of this investigation was to identify some of the talker-related acoustic-phonetic correlates of speech intelligibility. Specifically, we asked "What makes one talker more intelligible than another?" The results of this study showed that

global talker characteristics such as overall speaking rate and mean fundamental frequency did not correlate strongly with speech intelligibility scores. In contrast, we observed a tendency for a wider fundamental frequency range to be associated with higher overall intelligibility, and we found a significant gender-related difference in intelligibility such that the female talkers in our database were generally more intelligible than the male talkers. We also found strong evidence for a negative correlation between degree of vowel space reduction and overall intelligibility, suggesting that a talker's vowel space is a good indicator of overall speech intelligibility. Specifically, we found that talkers who produce vowels that are widely dispersed in the phonetic vowel space, particularly along the F1 dimension, have relatively high overall intelligibility scores. Finally, by examining two cases of common listener perceptual errors, segment deletion and mis-syllabification, we found that talkers with highly precise articulations at the fine-grained acoustic-phonetic level were less likely to be mis-heard than talkers who showed less articulatory precision.

These findings suggest that for normal talkers, although global characteristics do appear to have some bearing on overall intelligibility, there are substantially stronger correlations between fine-grained changes in articulation and speech intelligibility. In response to the question we posed at the start of this investigation, the present results suggest that highly intelligible talkers are those with a high degree of articulatory precision in producing segmental phonetic contrasts and a low degree of phonetic reduction in their speech. Based on these findings we can construct a profile of a highly intelligible talker: such a talker would be a female who produces sentences with a relatively wide range in fundamental frequency, employs a relatively expanded vowel space that covers a broad range in F1, precisely articulates her point vowels, and has a high precision of inter-segmental timing.

This characterization of a highly intelligible talker has several broader implications for our understanding of acoustic-phonetic variability and its effects on overall speech intelligibility. First, these findings suggest that a substantial portion of the observed variability in overall speech intelligibility can be traced directly to talker-specific characteristics of speech. As we noted in the introduction to this paper, the speech intelligibility scores in our database reflect both listener- and sentence-related characteristics as well as talker-related characteristics. Nevertheless, by focusing exclusively on talker-related characteristics, we were able to identify several correlates of variability in speech intelligibility.

A second implication of our findings deals with the role that talker-related characteristics play in situations that require "clear" speech. In a review of speech intelligibility tests used with disordered speakers, Weismer and Martin (1992) noted that indices of intelligibility deficits are considerably more useful when they include an explanatory component, in terms of the acoustic-phonetic bases of these deficits, that can serve as a guide for the remediation of such deficits. Knowledge of how speech varies across normal talkers, and how these variations affect speech intelligibility, might help direct attention to the crucial aspects of speech production for special populations, such as the hearing-impaired, and second language learners. Similarly, such fundamental knowledge about the production of normal speech may be very useful for the development of the next generation of speech output devices. For example, speech synthesizers and speech synthesis-by-rule systems could be designed to focus and emphasize the talker-characteristics that result in highly intelligible natural speech. Additionally, by knowing how natural speech varies across talkers and how these specific variations affect overall intelligibility, speech recognition systems might be able to achieve higher levels of performance over a much wider range of individual talkers and operational environments.

Finally, by establishing a direct link between overall speech intelligibility and some of the fine-grained acoustic-phonetic variations that exist across talkers, the results of this investigation add to the growing body of research demonstrating the important role that talker-specific attributes play in speech perception and spoken language processing. We believe it is now possible to provide a principled explanation for why some talkers are more intelligible than others and to specify the attributes of their speech with greater precision than has been possible in the past. Part of the success of this approach lies in having a large digital database of spoken sentences along with speech intelligibility scores for

each sentence. Thus, detailed acoustic-phonetic measures of the speech signal can be related directly to listeners' perceptual responses.

Acknowledgements

We are grateful to Luis Hernandez for technical support, to John Karl for compiling the Indiana Multi-talker Sentence Database, and to Christian Benoit for many useful comments. This research was supported by NIDCD Training Grant DC-00012 and by NIDCD Research Grant DC-00111 to Indiana University.

Appendix A. Signal captions

The following signals can be heard at the web site http//www.elsevier.nl/locate/specom.

Signal 1a. Audiofile of Talker F7's production of "It's easy to tell the depth of a well".

Signal 1b. Audiofile of Talker F7's production of "A pot of tea helps to pass the evening".

Signal 1c. Audiofile of Talker F7's production of "The horn of the car woke the sleeping cop".

Signal 2a. Audiofile of Talker M2's production of "It's easy to tell the depth of a well".

Signal 2b. Audiofile of Talker M2's production of "A pot of tea helps to pass the evening".

Signal 2c. Audiofile of Talker M2's production of "The horn of the car woke the sleeping cop".

Signal 3a. Audiofile of Talker M1's production of "The walled town was seized without a fight".

Signal 3b. Audiofile of Talker M9's production of "The walled town was seized without a fight".

Signal 4a. Audiofile of Talker F6's production of "The play seems dull and quite stupid".

Signal 4b. Audiofile of Talker F1's production of "The play seems dull and quite stupid".

References

J.W. Black (1957), "Multiple-choice intelligibility tests", *J. Speech and Hearing Disorders*, Vol. 22, pp. 213–235.

Z.S. Bond and T.J. Moore (1994), "A note on the acoustic-phonetic characteristics of inadvertently clear speech", *Speech Communication*, Vol. 14, No. 4, pp. 325–337.

A.R. Bradlow, L.C. Nygaard and D.B. Pisoni (1995), "On the contribution of instance-specific characteristics to speech perception", in: C. Sorin, J. Mariani, H. Meloni and J. Schoentgen, Eds., *Levels in Speech Communication: Relations and Interactions* (Elsevier, Amsterdam), pp. 13–25.

D. Byrd (1994), "Relations of sex and dialect to reduction", *Speech Communication*, Vol. 15, Nos. 1–2, pp. 39–54.

G. Fant (1973), *Speech Sounds and Features* (MIT Press, Cambridge, MA).

L.J. Gerstman (1968), "Classification of self-normalized vowels", *IEEE Trans. Audio Electroacoust.*, Vol. AU-16, pp. 78–80.

H.M. Hanson (1995), "Glottal characteristics of female speakers – Acoustic, physiological, and perceptual correlates", *J. Acoust. Soc. Amer.*, Vol. 97, No. 2, pp. 3422.

I.J. Hirsh, E.G. Reynolds and M. Joseph (1954), "Intelligibility of different speech materials", *J. Acoust. Soc. Amer.*, Vol. 26, pp. 530–538.

J.D. Hood and J.P. Poole (1980), "Influence of the speaker and other factors affecting speech intelligibility", *Audiology*, Vol. 19, pp. 434–455.

IEEE (1969), IEEE recommended practice for speech quality measurements, IEEE Report No. 297.

J. Karl and D. Pisoni (1994), "The role of talker-specific information in memory for spoken sentences", *J. Acoust. Soc. Amer.*, Vol. 95, p. 2873.

P.A. Keating, D. Byrd, E. Flemming and Y. Todaka (1994), "Phonetic analyses of word and segment variation using the TIMIT corpus of American English", *Speech Communication*, Vol. 14, No. 2, pp. 131–142.

D. Klatt and L. Klatt (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Amer.*, Vol. 87, pp. 820–857.

K.R. Kluender, R.L. Diehl and B.A. Wright (1988), "Vowel-length difference before voiced and voiceless consonants: An auditory explanation", *J. Phonetics*, Vol. 16, pp. 153–169.

J.C. Krause and L.D. Braida (1995), "The effects of speaking rate on the intelligibility of speech for various speaking modes", *J. Acoust. Soc. Amer.*, Vol. 98, No. 2, pp. 2982.

P. Ladefoged and D.E. Broadbent (1957), "Information conveyed by vowels", *J. Acoust. Soc. Amer.*, Vol. 29, pp. 98–104.

L. Lamel, R. Kassel and S. Seneff (1986), "Speech database development: Design and analysis of the acoustic-phonetic corpus", *Proc. DARPA Speech Recognition Workshop*, February 1986, pp. 100–109.

J. Laver and P. Trudgill (1979), "Phonetic and linguistic markers in speech", in: K.R. Scherer and H. Giles, Eds., *Social Markers in Speech* (Cambridge University Press, Cambridge), pp. 1–32.

B. Lindblom (1990), "Explaining phonetic variation: A sketch of the H & H theory", in: W.J. Hardcastle and A. Marchal, Eds., *Speech Production and Speech Modeling* (Kluwer Academic Publishers, Dordrecht), pp. 403–439.

P.A. Luce and T.D. Carrell (1981), Creating and editing waveforms using WAVES, Research in Speech Perception Progress Report No. 7 (Indiana University Speech Research Laboratory, Bloomington).

J.L. Miller (1981), "Effects of speaking rate on segmental distinc-

tions", in: P.D. Eimas and J.L. Miller, Eds., *Perspectives on the Study of Speech* (Lawrence Erlbaum, Hillsdale, NJ), pp. 39–74.

R.B. Monsen (1976), "Normal and reduced phonological space: the productions of English vowels by deaf adolescents", *J. Phonetics*, Vol. 4, pp. 189–198.

S.-J. Moon and B. Lindblom (1994), "Interaction between duration, context and speaking style in English stressed vowels", *J. Acoust. Soc. Amer.*, Vol. 96, pp. 40–55.

J.W. Mullennix, D.B. Pisoni and C.S. Martin (1989), "Some effects of talker variability on spoken word recognition", *J. Acoust. Soc. Amer.*, Vol. 85, pp. 365–378.

A.T. Neel (1995), "Intelligibility of normal speakers: Error analysis", *J. Acoust. Soc. Amer.*, Vol. 98, p. 2982.

L.C. Nygaard, M.S. Sommers and D.B. Pisoni (1994), "Speech perception as a talker-contingent process", *Psychological Sci.*, Vol. 5, pp. 42–46.

L.C. Nygaard, M.S. Sommers and D.B. Pisoni (1995), "Effects of stimulus variability on perception and representation of spoken words in memory", *Perception and Psychophysics*, Vol. 57, pp. 989–1001.

D. Pallett (1990), "Speech corpora and performance assessment in the DARPA SLS program", *Proc. Internat. Conf. on Spoken Language Processing 1990*, pp. 24.3.1–24.3.4.

T.J. Palmeri, S.D. Goldinger and D.B. Pisoni (1993), "Episodic encoding of voice attributes and recognition memory for spoken words", *J. Experimental Psychology: Learning, Memory and Cognition*, Vol. 19, pp. 1–20.

E.M. Parker, R.L. Diehl and K.R. Kluender (1986), "Trading relations in speech and nonspeech", *Perception and Psychophysics*, Vol. 34, pp. 314–322.

M.A. Picheny, N.I. Durlach and L.D. Braida (1985), "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech", *J. Speech and Hearing Research*, Vol. 28, pp. 96–103.

M.A. Picheny, N.I. Durlach and L.D. Braida (1986), "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech", *J. Speech and Hearing Research*, Vol. 29, pp. 434–446.

M.A. Picheny, N.I. Durlach and L.D. Braida (1989), "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech", *J. Speech and Hearing Research*, Vol. 32, pp. 600–603.

D.B. Pisoni (1993), "Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning", *Speech Communication*, Vol. 13, Nos. 1–2, pp. 109–125.

R.F. Port (1981), "Linguistic timing factors in combination", *J. Acoust. Soc. Amer.*, Vol. 69, pp. 262–274.

R.F. Port and J. Dalby (1982), "Consonant/vowel ratio as a cue for voicing in English", *Perception and Psychophysics*, Vol. 32, pp. 141–152.

R.P. Runyon and A. Haber (1991), *Fundamentals of Behavioral Statistics* (McGraw-Hill, New York), pp. 201–205.

M.S. Sommers, L.C. Nygaard and D.B. Pisoni (1994), "Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude", *J. Acoust. Soc. Amer.*, Vol. 96, pp. 1314–1324.

M.T.J. Tielen (1992), Male and Female Speech: An experimental study of sex-related voice and pronunciation characteristics, Doctoral dissertation, University of Amsterdam.

R.M. Uchanski, S. Choi, L.D. Braida, C.M. Reed and N.I. Durlach (1996), "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate", *J. Speech and Hearing Research*, Vol. 39, pp. 494–509.

G. Weismer and R.E. Martin (1992), "Acoustic and perceptual approaches to the study of intelligibility", in: R.D. Kent, Ed., *Intelligibility in Speech Disorders: Theory, Measurement and Management* (Benjamins, Amsterdam/Philadelphia), pp. 67–118.

V. Zue, S. Seneff and J. Glass (1990), "Speech database development at MIT: TIMIT and beyond", *Speech Communication*, Vol. 9, No. 4. pp. 351–356.