УДК 004.934.1 DOI: 10.18698/2541-8009-2019-02-436

СРАВНЕНИЕ РАЗЛИЧНЫХ КРИТЕРИЕВ СЕГМЕНТАЦИИ ДЛЯ ВЫДЕЛЕНИЯ УЧАСТКОВ РЕЧЕВЫХ СИГНАЛОВ ВО ВРЕМЕННОЙ ОБЛАСТИ

А.Б. Жукова zhukovaab@student.bmstu.ru
SPIN-код: 1923-4866

А.Л. Масленников amas@bmstu.ru
SPIN-код: 7087-6303

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Распознавание речи — сложная техническая задача, над решением которой работают многие ученые и коммерческие компании. Решение этой задачи во временной области, как правило, требует предварительной сегментации речевого сигнала, т. е. выделения участков речевых сигналов, соответствующих словам, слогам или фонемам. Для этого вводят критерии сегментации, связанные с мощностью или с частотой изменения речевого сигнала на некотором интервале времени. Критерии сегментации могут быть сформулированы по-разному, что отражается в первую очередь на сложности их алгоритмической реализации и объеме требуемых вычислительных ресурсов. В статье выполнено сравнение нескольких критериев сегментации при выделении участков речевых сигналов, соответствующих отдельным словам.

Ключевые слова

Распознавание речи, голосовое управление, сегментация речевого сигнала, критерий сегментации, фильтр Савицкого — Голея, фильтр скользящего среднего, скользящая дисперсия, выделение участков речевых сигналов

Поступила в редакцию 17.01.2019

Введение. Задача распознавания речи продолжает быть актуальной и по сей день. Ее решение необходимо, например, для формирования систем с голосовым управлением, распознавания больших объемов текста или идентификации личности по голосу [1–4]. Существующие методы распознавания речи могут быть как временными, так и частотными. В общем случае наилучший результат по точности распознавания речи можно получить путем комбинации временных и частотных методов [5]. Распознавание речи во временной области зачастую включает в себя сегментацию речевого сигнала: выделение участков со словами, слогами или фонемами, а затем уже определение характеристических параметров голоса или непосредственно преобразование речевого сигнала в текстовую информацию [6, 7].

Для сегментации речевого сигнала необходимо задание критерия сегментации, как правило, связанного с уровнем мощности речевого сигнала на некотором интервале времени. В этом случае индикатором выделяемого участка является пре-

вышение значения выбранного критерия сегментации заданного порогового значения на некотором интервале времени, что проиллюстрировано на рис. 1.

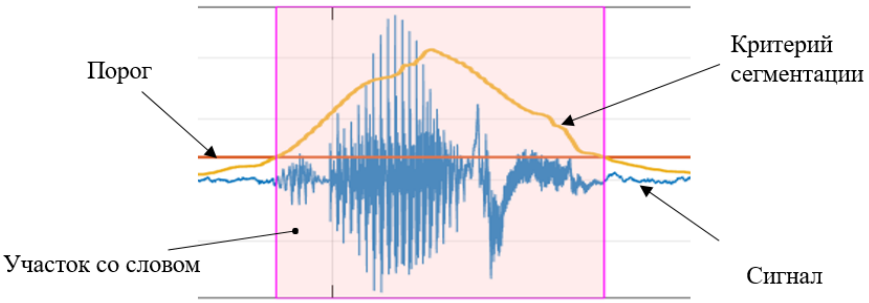


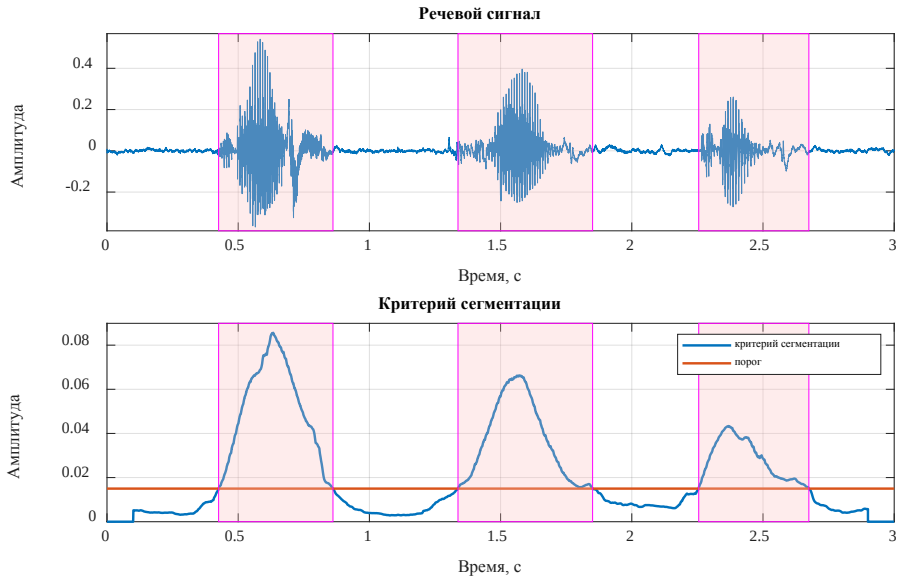
Рис. 1. Выделение участка речевого сигнала

Критерии сегментации. В идеальном случае, когда при проведении эксперимента исключается запись одновременной речи нескольких человек, а также сводятся к минимуму акустические помехи и шумы, можно рассматривать сегментацию речевого сигнала как выделение непосредственно слов и слогов. В данной работе осуществляется сравнение нескольких различных критериев сегментации, используемых для решения рассматриваемой задачи во временной области, а именно:

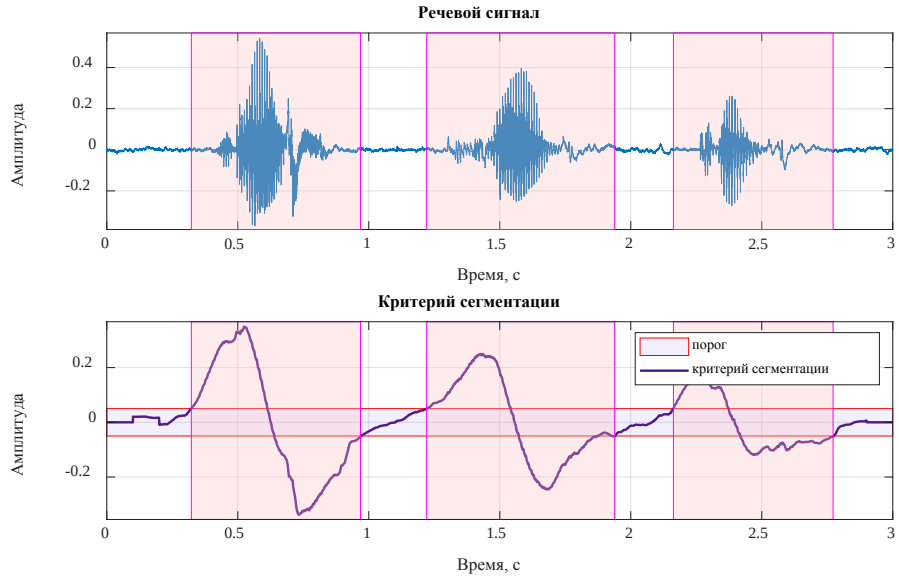
– порог огибающей речевого сигнала (с фильтром скользящего среднего);
– порог огибающей производной речевого сигнала (с фильтром скользящего среднего);
– порог скользящей дисперсии речевого сигнала;
– порог огибающей речевого сигнала (с фильтром Савицкого — Голея);
– порог огибающей производной речевого сигнала (с фильтром Савицкого — Голея).

Для сравнения перечисленных выше критериев сегментации использовали записанный речевой сигнал, в котором явно выделяются три участка со словами «раз», «два» и «три», как показано на рис. 2, *а*. Отметим, что применение фильтров при формировании критериев сегментации необходимо для уменьшения влияния случайной составляющей и, как следствие, к снижению количества ложных срабатываний алгоритма выделения участка.

Порог огибающей речевого сигнала (с фильтром скользящего среднего). Для формирования критерия сегментации сначала для полученного речевого сигнала вычисляют его модуль (формируют огибающую), а затем применяют фильтр скользящего среднего. Превышение значения огибающей речевого сигнала заданного порогового значения является индикатором слова, как проиллюстрировано на рис. 2. Отметим, что скользящее среднее будет смещено по времени, поэтому полученная огибающая должна быть сдвинута по времени на половину точек усреднения. Метод сегментации на базе данного критерия сегментации достаточно прост, однако для его реализации требуется нетривиальный подбор порогового значения, а также выбора оптимального количества точек усреднения.



а



б

Рис. 2. Пример сегментации речевого сигнала с критерием сегментации по порогу:

а — огибающей сигнала; *б* — огибающей производной

Порог огибающей производной речевого сигнала (с фильтром скользящего среднего). Помимо огибающей речевого сигнала можно использовать огибающую производной речевого сигнала. Пример использования этого критерия сегментации проиллюстрирован на рис. 2, *б*. Увеличение значения производной происходит при увеличении амплитуды речевого сигнала, т. е. при увеличении мощности сигнала на небольшом интервале времени. Как и в предыдущем случае, к численно рассчи-

тываемой производной применяется фильтр скользящего среднего. Индикатором слова является участок речевого сигнала, где значение огибающей производной не просто превышает значение заданного порога, а выходит за пределы заданной трубки точности. Для сегментации на базе этого критерия сегментации требуется алгоритмическая компенсация частых ложных срабатываний алгоритма, но при этом выбор порогового значения (ширина трубки) более прост.

Порог скользящей дисперсии речевого сигнала. По аналогии с вычислением производной, значение которой для речевых сигналов характеризует меру разброса амплитуды относительно нуля, можно использовать в качестве критерия сегментации скользящую дисперсию. На участке, соответствующему слову, дисперсия сигнала увеличивается, а на интервале с паузой — стремится к нулю (точнее говоря, к суммарному уровню шумов), что проиллюстрировано на рис. 3. По сравнению с использованием огибающей производной, для использования скользящей дисперсии не требуется большое число алгоритмических компенсаций ложных срабатываний алгоритма, поскольку скользящая дисперсия является более гладкой функцией. Задание порогового значения в данном случае можно связать с уровнем измерительных и акустических шумов.

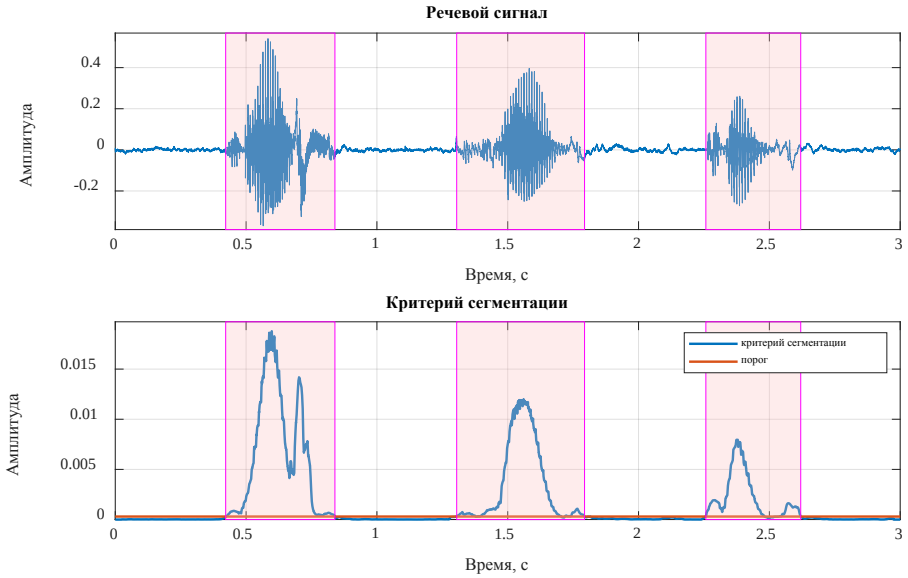
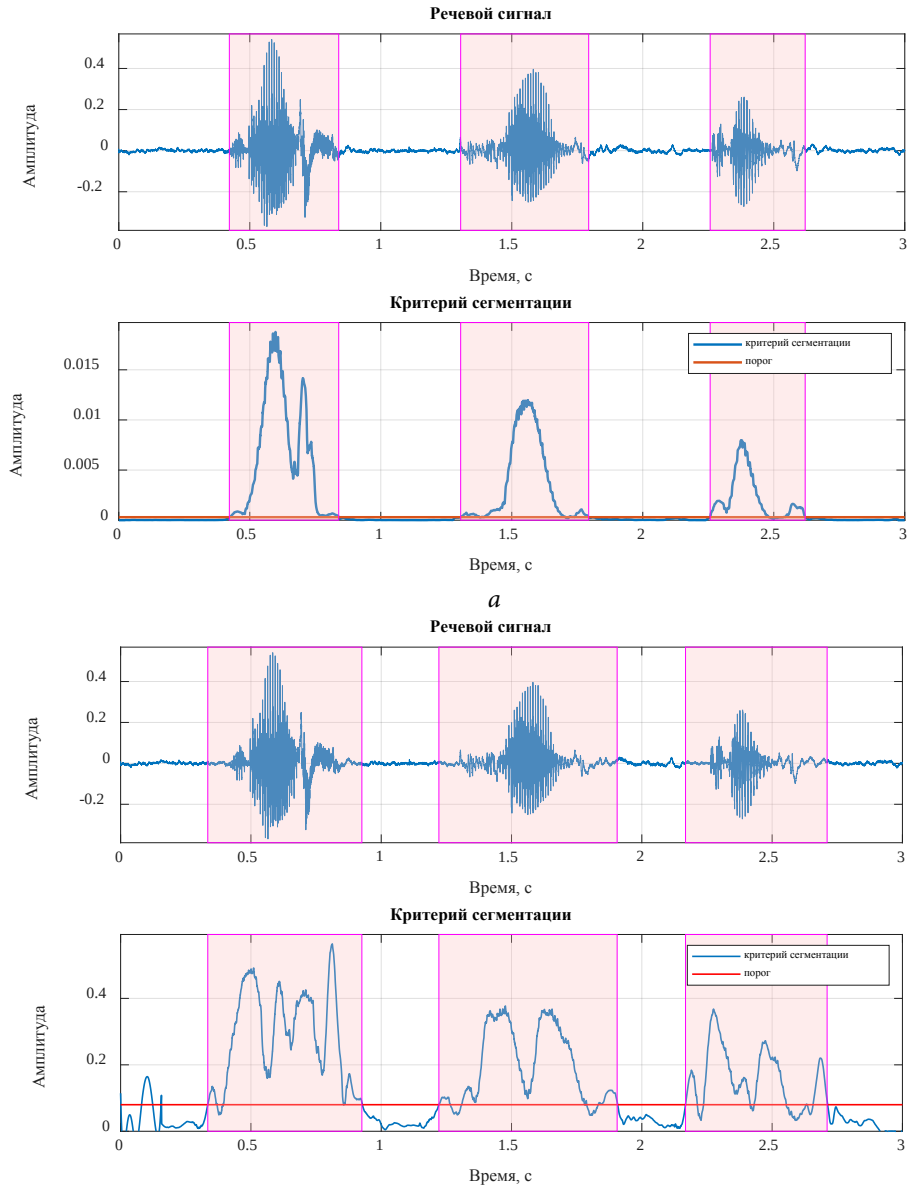


Рис. 3. Пример сегментации речевого сигнала с критерием сегментации по скользящей дисперсии

Порог огибающей речевого сигнала и его производной (с фильтром Савицкого — Голея). Очевидным недостатком используемой в предыдущих индикаторах фильтрации скользящим средним является наличие смещения по времени (запаздывания) критерия сегментации относительно исходного речевого сигнала, а также сложность выбора количества точек усреднения. В качестве альтернативы скользящему среднему для формирования огибающей можно использовать фильтр Савицкого — Голея (который оптимален в контексте минимизации

среднеквадратичной погрешности) [8–10]. Применение этого фильтра не приводит к появлению запаздывания между критерием сегментации и речевым сигналом, что проиллюстрировано для огибающей речевого сигнала на рис. 4, *а*, а для огибающей производной речевого сигнала — на рис. 4, *б*. Недостатком использования фильтра Савицкого — Голея является более сложный для реализации алгоритм фильтрации, который требует дополнительных вычислений.



а

б

Рис. 4. Пример сегментации речевого сигнала с критерием сегментации (с применением фильтра Савицкого — Голея) по порогу:

а — огибающей сигнала; *б* — огибающей производной

Сравнение качества выделения участков речевых сигналов различными индикаторами. Для сравнения качества сегментации речевых сигналов для описанных выше критериев сегментации использовался речевой сигнал, содержащий 32 слова. В результате его сегментации определяли ошибку распознавания количества слов с использованием каждого критерия сегментации. Значения полученных ошибок приведены в таблице, где отчетливо видно, что использование критериев сегментации по порогу огибающей производной с применением фильтра скользящего среднего и по порогу огибающей сигнала с применением фильтра Савицкого — Голея имеют весьма большую ошибку. Остальные три индикатора имеют небольшую ошибку, что делает их более пригодными для практических задач. Отметим, что, несмотря на небольшое значение ошибки для критерия сегментации по порогу огибающей речевого сигнала с фильтром Савицкого — Голея, его практическая реализация сложнее, особенно в системах реального времени, где объем вычислений должен сводиться к минимуму.

Сравнение критериев сегментации

Критерий сегментации	Процент ошибок
Порог огибающей речевого сигнала (фильтр скользящего средним)	3
Порог огибающей производной речевого сигнала (фильтр скользящего средним)	53
Порог скользящей дисперсии речевого сигнала	0
Порог огибающей речевого сигнала (фильтр Савицкого — Голея)	21
Порог огибающей производной речевого сигнала (фильтр Савицкого — Голея)	9

Выводы. В данной статье рассмотрены различные критерии сегментации, с помощью которых можно осуществлять выделение участков речевых сигналов, соответствующих словам во временной области. Сравнение критериев сегментации осуществлялось по простоте реализации метода сегментации на их базе, сложности выбора порогового значения и необходимости использования дополнительных вычислений.

Наилучшим критерием сегментации можно считать критерий сегментации по порогу скользящей дисперсии речевого сигнала. Сложность его реализации мала по сравнению с критериями сегментации, в которых используется фильтр Савицкого — Голея, и соизмерима с остальными, однако выбор порогового значения существенно проще, так как его можно связать с дисперсией измерительного шума микрофона и акустического шума в помещении.

Литература

[1] Плотников В.Н., Суханов В.А., Жигулевцев Ю.Н. Речевой диалог в системах управления. М., Машиностроение, 1988.

[2] Рабинер Л.Р., Шафер Р.В. Цифровая обработка сигналов. М., Радио и связь, 1981.

[3] Сапожков М.А. Речевой сигнал в кибернетике и связи. М., Связьиздат, 1963.
[4] Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. Киев, Наукова думка, 1987.
[5] Алимурадов А.К., Чураков П.П. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи. *Измерение. Мониторинг. Управление. Контроль*, 2015, № 2(12), с. 27–35.
[6] Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. *Информационные процессы*, 2004, т. 4, № 2, с. 202–220.
[7] Цыплихин А.И., Сорокин В.Н. Сегментация речи на кардинальные элементы. *Информационные процессы*, 2006, т. 6, № 3, с. 177–207.
[8] Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 1964, vol. 36, no. 8, pp. 1627–1639. DOI: 10.1021/ac60214a047 URL: https://pubs.acs.org/doi/10.1021/ac60214a047
[9] Savitzky A. A historic collaboration. *Anal. Chem.*, 1989, vol. 61, no. 15, pp. 921A–923A. DOI: 10.1021/ac00190a003 URL: https://pubs.acs.org/doi/10.1021/ac00190a003
[10] Steinier J., Termonia Y., Deltour J. Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.*, 1972, vol. 44, no. 11, pp. 1906–1909. DOI: 10.1021/ac60319a045 URL: https://pubs.acs.org/doi/10.1021/ac60319a045

Жукова Анастасия Борисовна — студентка кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Масленников Андрей Леонидович — ассистент кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Задорожная Наталия Михайловна, кандидат технических наук, доцент кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

DIFFERENT SEGMENTATION CRITERIA COMPARISON IN TIME-DOMAIN SPEECH SEGMENTATION PROBLEM

A.B. Zhukova zhukovaab@student.bmstu.ru
SPIN-code: 1923-4866

A.L. Maslennikov amas@bmstu.ru
SPIN-code: 7087-6303

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

Speech recognition is a complex technical problem which is of interest of many scientists and commercial companies. Its solution in time-domain typically requires preliminary speech segmentation, i.e. extraction of words, syllables or phonemes. In order to do that different segmentation criteria are used. Typically, those criteria are associated with signal power or signal changes frequency during a specified time interval. Segmentation criteria could be formulated differently, that results in different algorithmically and computational complexity. In this paper different segmentation criteria (associated with signal power) for extracting words from a speech signal are comparing.

Keywords

Speech recognition, voice control, speech segmentation, segmentation criteria, Savitzky-Golay filter, moving-average filter, moving-variance filter, speech segmentation

Received 17.01.2019

References

[1] Plotnikov V.N., Sukhanov V.A., Zhigulevtsev Yu.N. Rechevoy dialog v sistemakh upravleniya [Speech dialogue in control systems]. Moscow, Mashinostroenie Publ., 1988 (in Russ.).

[2] Rabiner L.R., Schafer R.W. Digital processing of speech signals. Pearson, 1978. (Russ. ed.: Tsifrovaya obrabotka signalov. Moscow, Radio i svyaz' Publ., 1981.)

[3] Sapozhkov M.A. Rechevoy signal v kibernetike i svyazi [Speech signal in cybernetics and communication]. Moscow, Svyaz'izdat Publ., 1963 (in Russ.).

[4] Vintsyuk T.K. Analiz, raspoznavanie i interpretatsiya rechevykh signalov [Analysis, recognition and interpretation of speech signals]. Kiev, Naukova dumka Publ., 1987 (in Russ.).

[5] Alimuradov A.K., Churakov P.P. Review and classification methods for processing speech signals in the speech recognition systems. *Izmerenie. Monitoring. Upravlenie. Kontrol'* [Measuring. Monitoring. Management. Control], 2015, no. 2(12), pp. 27–35 (in Russ.).

[6] Sorokin V.N., Tsyplikhin A.I. Segmentation and recognition of vowels. *Informatsionnye protsessy* [Information Processes], 2004, vol. 4, no. 2, pp. 202–220 (in Russ.).

[7] Tsyplikhin A.I., Sorokin V.N. Speech segmentation into principal elements. *Informatsionnye protsessy* [Information Processes], 2006, vol. 6, no. 3, pp. 177–207 (in Russ.).

[8] Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 1964, vol. 36, no. 8, pp. 1627–1639. DOI: 10.1021/ac60214a047 URL: https://pubs.acs.org/doi/10.1021/ac60214a047

[9] Savitzky A. A historic collaboration. *Anal. Chem.*, 1989, vol. 61, no. 15, pp. 921A–923A. DOI: 10.1021/ac00190a003 URL: https://pubs.acs.org/doi/10.1021/ac00190a003

[10] Steinier J., Termonia Y., Deltour J. Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.*, 1972, vol. 44, no. 11, pp. 1906–1909. DOI: 10.1021/ac60319a045 URL: https://pubs.acs.org/doi/10.1021/ac60319a045

Zhukova A.B. — Student, Department of Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Maslennikov A.L. — Teaching Assistant, Department of Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Zadorozhnaia N.M., Cand. Sc. (Eng.), Assoc. Professor, Department of Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.