NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

PROJECT TITLE GOES HERE

YOUR NAME
Matriculation number

Supervisor: XXXXX
Co-supervisor: XXXXX

A Final Year Report submitted to Asian School of the Environment, Nanyang Technological University in partial fulfilment of the requirements for the Degree of

BACHELOR OF SCIENCE WITH HONOURS IN ENVIRONMENTAL EARTH SYSTEMS SCIENCE

2020/2021 Semester 1

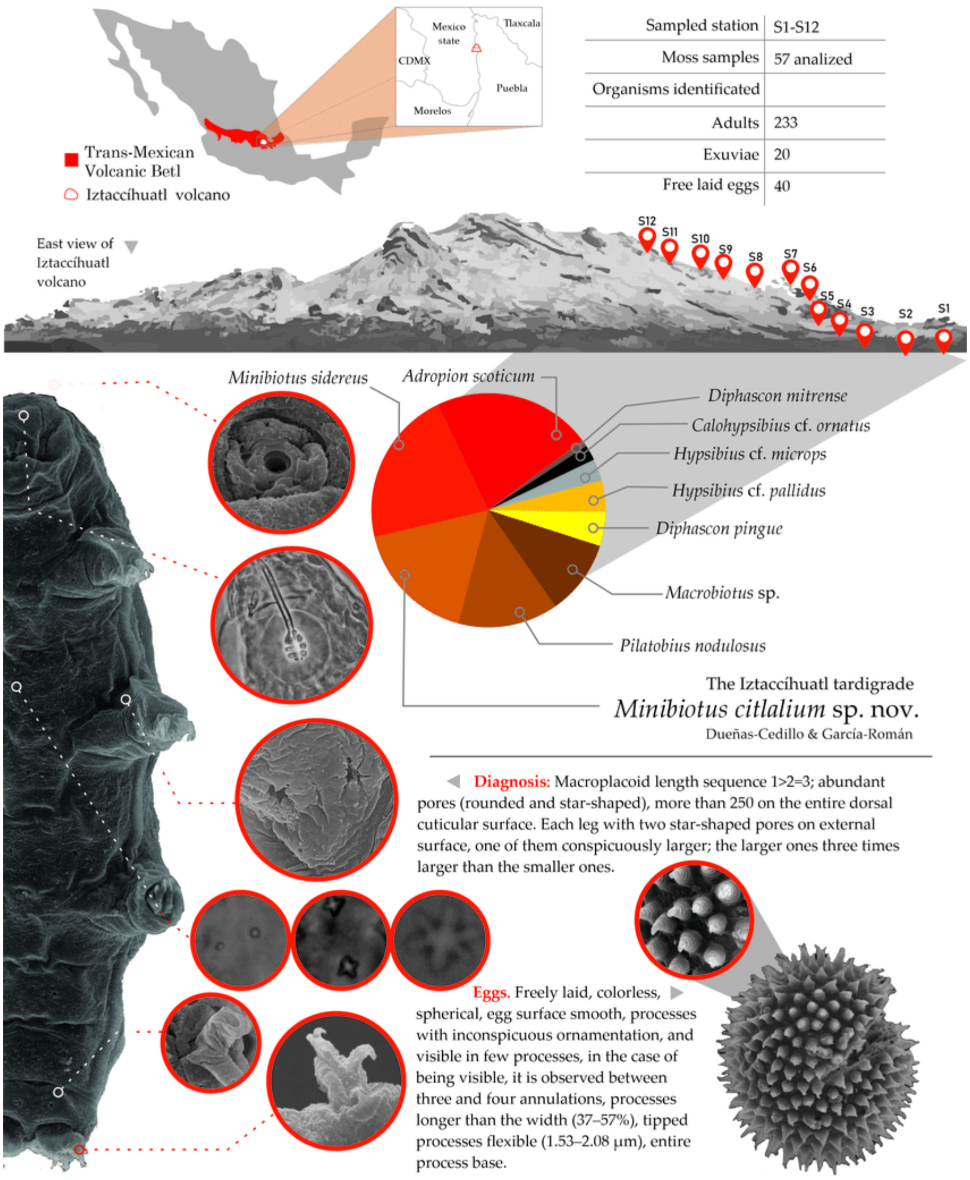
Abstract

Vivamus non vehicula tortor. Aenean tincidunt vel nisi id tincidunt. Aliquam eget sollicitudin ipsum. Curabitur quis malesuada magna. Integer nec condimentum sapien. Donec finibus rutrum tellus vel feugiat. Fusce auctor lectus id nunc dignissim, et ultricies eros volutpat. Pellentesque condimentum sed mi et placerat. Morbi semper, erat vitae sodales lobortis, orci enim viverra urna, et mollis mi dui ut ante. Vestibulum sodales lacinia scelerisque. Donec ut augue tellus. Etiam sit amet lacinia ipsum. Morbi commodo, sem sollicitudin dictum fringilla, libero quam varius erat, vel elementum felis velit ut elit. Fusce at sagittis eros. Phasellus tempor ullamcorper massa, consectetur scelerisque purus facilisis in. Vestibulum tempus eros vel sapien rutrum ullamcorper. Fusce at facilisis risus. Donec porttitor augue eu quam tempor, eget cursus risus commodo. Pellentesque sollicitudin, ipsum in tincidunt sagittis, nibh metus ultrices eros, egestas iaculis arcu justo non lectus. Praesent mollis orci ac sodales mollis. Sed suscipit sollicitudin ex id volutpat. Praesent laoreet molestie cursus. Nam dapibus dapibus bibendum. Donec maximus, velit non commodo porttitor, eros lacus ultrices ante, nec interdum purus mauris quis turpis. Nam dapibus quis lacus sed finibus. Aliquam venenatis, ex sit amet suscipit tempor, leo metus porta nulla, a egestas dui mi non risus. Duis vel libero ut mauris sodales cursus nec nec eros. Donec accumsan justo dolor. Quisque varius sem id dui dignissim, sed porta massa sodales. Nulla facilisi. Phasellus id metus ut enim eleifend volutpat. Fusce iaculis sodales tincidunt. Fusce quis convallis justo. Phasellus elementum, felis vel ultrices ultricies, ex nulla euismod justo, et.

Keywords: Blind source separation, Independent component analysis, Independent vector analysis, Sparse component analysis, Automatic speech recognition, Android application development.

Graphical Abstract

Proin semper mauris arcu, at sodales dolor lobortis eu. Vestibulum non tempus nibh. Proin id mauris tincidunt, hendrerit ipsum at, placerat lacus. Proin ac malesuada mauris. Ut justo arcu, euismod ac magna sit amet, rutrum gravida sapien. Vestibulum efficitur sodales mattis. Mauris ornare leo nec elit fermentum scelerisque. Nullam nec augue sed ex fringilla sagittis. Suspendisse tincidunt sollicitudin tortor eget vehicula. Sed volutpat scelerisque augue sit amet sodales. Donec eget nisi facilisis, viverra magna ac, cursus purus. Mauris eget augue sit amet lectus tempus lacinia eget eu nisi. Morbi rutrum eros non sapien euismod, id sagittis ipsum sagittis. Phasellus eget.



Acknowledgement

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Table of Contents

Lists of Figures

List of Tables

1. Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sit amet lacinia urna, eu posuere enim. Donec vulputate finibus molestie. Praesent sit amet lacus orci. Etiam molestie augue a ligula ullamcorper, lacinia lobortis elit semper. Pellentesque a tincidunt libero. Morbi nec volutpat sapien, at pretium diam. Nunc porttitor porttitor efficitur. Maecenas lobortis feugiat eros. Integer ac tortor a nulla porta scelerisque. Suspendisse tempus sem non lectus malesuada, nec elementum orci vestibulum. Donec mattis, nunc vel ornare facilisis, metus libero volutpat ex, eu sodales ex urna nec nibh. Donec rhoncus finibus ultricies. Cras pretium libero a scelerisque placerat. Suspendisse ac ante in velit sagittis gravida vehicula vel diam. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

Praesent eget accumsan tellus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris eu sem ac ex ultricies sollicitudin eget quis leo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Morbi sed ligula diam. Nullam vel suscipit libero. Donec ut felis a quam porttitor feugiat id id odio. Sed feugiat ex et mauris malesuada pretium. Mauris consequat, tellus nec lacinia condimentum, quam diam viverra justo, in vestibulum magna quam vitae sem. Nam varius, lacus sed scelerisque tincidunt, nunc nunc laoreet nisl, eu cursus lorem nisi a augue. Nullam pellentesque id mauris eget lobortis. Donec eget arcu suscipit, vestibulum metus et, fringilla tortor. Praesent semper interdum nisi sed consectetur. Sed tristique venenatis odio, sed sollicitudin sem placerat in (Smith III, 2011).

Praesent blandit suscipit ex eget sodales. Integer commodo aliquam accumsan. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque lobortis porttitor auctor. Nulla magna velit, sagittis sollicitudin mattis ac, sagittis vitae risus. Donec quis varius dui. Curabitur rutrum lacus id dui molestie rutrum. Cras ut felis accumsan, porta risus nec, cursus lorem. Nullam aliquet mollis dui, id tincidunt quam tempor eu. Aliquam volutpat leo id facilisis porttitor. Phasellus eleifend mauris vel arcu suscipit volutpat. Sed lacinia ipsum nunc, sit amet elementum turpis condimentum at. Aenean a condimentum orci, at vehicula velit. Aliquam venenatis dolor in erat tincidunt, a elementum dui aliquet. Etiam gravida accumsan erat, at consectetur enim auctor eu. Integer eget dui et enim euismod placerat ut et augue.

Etiam nec tincidunt risus, eu fermentum nulla. Fusce pulvinar urna ac nunc imperdiet, porta mollis odio aliquam. Nam at rutrum enim. Vestibulum odio sapien, aliquet et venenatis suscipit, molestie ac neque. Maecenas mollis eu est et euismod. Donec mattis, ipsum a blandit dignissim, massa turpis cursus nisi, id rutrum diam diam vitae magna. Phasellus et metus nisl. Fusce ac tellus ac turpis blandit rutrum nec eget justo. Morbi a mauris at nunc scelerisque blandit. Pellentesque eu ex dapibus, pellentesque lorem sit amet, commodo tortor. Nunc ornare libero ut faucibus faucibus. Etiam lectus velit, sodales in sem in, scelerisque venenatis ex (Scheibler, 2018). Pellentesque sit amet porttitor ante, at sodales ex. Nulla volutpat diam ac justo vehicula, eget vestibulum ex dapibus.

Nulla at luctus augue. Suspendisse potenti. Vestibulum condimentum nisl lectus, et porta erat efficitur et. Curabitur diam ligula, efficitur nec feugiat a, eleifend a eros. Sed a magna ac dolor sagittis bibendum sit amet id mi. Phasellus id mi sed mi lobortis aliquet ut quis nulla. Sed consequat mi vel libero condimentum, faucibus volutpat velit porttitor. In risus lacus, porttitor ac diam vestibulum, mattis euismod eros. Phasellus auctor iaculis fermentum. Fusce nec sollicitudin nulla, eget porta orci. Integer fringilla in massa quis eleifend.

Cras dolor arcu, auctor in tristique nec, faucibus sed libero. Nulla rutrum malesuada viverra. Vivamus ut consequat arcu, maximus pulvinar eros. Sed faucibus posuere magna sed elementum. Curabitur eu varius lorem. Maecenas pharetra dolor non purus bibendum interdum. Quisque in enim justo. Proin accumsan, magna ac sollicitudin vehicula, nulla ligula tempor nunc, id consequat erat nulla eget risus. Nullam ullamcorper suscipit bibendum. Nam molestie laoreet ipsum, at lacinia lorem dictum in. Pellentesque lacinia et dolor et condimentum. Curabitur ac imperdiet neque, in dictum est. Integer sollicitudin, sapien sed ultricies maximus, metus leo sodales sem, vel dignissim nisi odio vitae quam. Suspendisse quam nibh, accumsan ac eros at, lacinia eleifend nunc. Quisque tincidunt metus dolor, nec tristique quam fermentum aliquet. Sed vehicula sed est id mattis.

Donec id rutrum quam. Quisque quis magna et elit tempus interdum sit amet eu ante. Maecenas eget dolor tempus, sollicitudin velit id, auctor risus. Pellentesque nec felis ex. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nunc urna dolor, molestie sed ullamcorper at, mollis eu tortor. Praesent ullamcorper leo id metus egestas, a finibus nunc finibus. Nunc mattis mauris libero, vitae faucibus eros efficitur in. Quisque quis ex consequat, tempus turpis non, volutpat justo. Pellentesque consequat elit ac mattis porttitor.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec sodales luctus elementum. Quisque sit amet dolor at metus convallis tincidunt. Curabitur id quam pharetra, suscipit nulla a, gravida leo. Nulla sodales nulla ac nibh ornare volutpat. Cras vitae ullamcorper mauris, quis egestas lacus. Proin lacinia malesuada laoreet. Praesent vel elit hendrerit, congue lorem et, placerat odio. Pellentesque consectetur rutrum ipsum quis dignissim.

This project aims to develop an Android application, implementing blind source separation to improve the performance of automatic speech recognition for recordings with multiple speech sources.

2. Material and Methods

2.1 Data Collection

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sit amet lacinia urna, eu posuere enim. Donec vulputate finibus molestie. Praesent sit amet lacus orci. Etiam molestie augue a ligula ullamcorper, lacinia lobortis elit semper in Fig 2-1.

Figure 2-1:

Adapted from Sapp, 2004

2.2 Maecenas lobortis

The summary of speaker information for the evaluation sources and analyses of **pulse code modulation (PCM)** used in this report is presented in Table A-1.

2.2.1 Frequency-domain ipsum dolor

The various types of window functions and their properties are reviewed in Harrris et al.(1978). For this project, we use the periodic Hamming window defined by

$$w(\tau) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi\tau}{R}\right) & 0 \leq \tau < R, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

with 75% overlap.

2.3 Vestibulum tempus

A set of N source signals $\boldsymbol{s}(\tau) = [s_0(\tau), \ldots, s_{N-1}(\tau)]^T \in \mathbb{R}^{N \times T_t}$ over the samples $0 \leq \tau < T_t$, such that $s_n = [s_n(\tau = 0), s_n(1), \ldots, s_n(T_t - 1)]$ for $0 \leq n < N$, undergoes mixing whose result is observed by an array of M sensors (in this project's context, microphones). The observed signals are represented by $\boldsymbol{x}(\tau) = [x_0(\tau), \ldots, x_{M-1}(\tau)]^T$, such that $x_m = [x_m(\tau = 0), x_m(1), \ldots, x_m(T_t - 1)] \in \mathbb{R}^{M \times T}$ for $0 \leq m < M$. The sources are mixed convolutively such that

$$\boldsymbol{x}(\tau) = \sum_{k=0}^{K-1} \boldsymbol{A}_k \boldsymbol{s}(\tau - k) \tag{2.2}$$

where $\boldsymbol{A}_k \in \mathbb{R}^{M \times N}$ with K often assumed, in practice, to be some finite positive integer. This means that $\boldsymbol{x}(\tau)$ can be thought of as the output signal of a finite impulse response filter with input $\boldsymbol{s}(\tau)$ and $\boldsymbol{A}_k$ representing the coefficients of the kth-order filter.

However, solving a convolutive system in the time domain is difficult and requires significant computational power. To simplify the problem, BSS problems are often transformed into the frequency domain such that

$$\boldsymbol{X}(\omega, t) = \boldsymbol{A}(\omega) \boldsymbol{S}(\omega, t) \tag{2.3}$$

where $\boldsymbol{X}(\omega, t) \in \mathbb{C}^{M \times T_f}$ and $\boldsymbol{S}(\omega, t) \in \mathbb{C}^{N \times T_f}$ are the frequency-domain representation of the observed signals and the source signals respectively (Smaragdis, 1998).

The configuration of 2 sources used is detailed in Figures A-2 to A-5.

3. Results and Discussion

3.1 Main Activity

A crucial part of an Android application comprises of ‘activities’. A set of transcript evaluation were assumed in this project and are summarized at Table A-2

Table 3-1 shows the summary of the performance for each algorithm by acoustic environment (Env.) and the average runtimes (Rt.) for two-source separation.

	Algorithm	Rt. (s)	Env.	SDR	SIR	SAR	STOI	ESTOI	WER
Raw	AuxIVA	358.2	LTH	7.218	8.883	**13.310**	0.675	0.503	0.392
			OFC	**8.240**	10.302	**13.316**	0.705	0.544	0.310
			OTD	**14.357**	17.246	**18.263**	0.780	0.611	0.161
	FastICA	135.2	LTH	**9.564**	**15.960**	10.961	0.778	**0.609**	**0.155**
			OFC	7.873	**13.206**	10.001	**0.760**	**0.599**	**0.207**
			OTD	10.421	**18.700**	11.281	0.781	0.615	**0.146**
	SCA-DSF	**96.1**	LTH	9.425	15.675	10.917	**0.779**	0.606	0.167
			OFC	7.300	12.221	9.805	0.754	0.595	0.224
			OTD	9.705	17.472	10.852	**0.785**	**0.622**	0.162
Improvement	AuxIVA		LTH	16.959	8.838	19.726	0.423	0.330	-0.587
			OFC	16.471	10.258	16.212	0.394	0.315	-0.690
			OTD	23.727	17.041	24.701	0.483	0.440	-0.820
	FastICA		LTH	19.304	15.915	17.377	0.527	0.436	-0.824
			OFC	16.104	13.161	12.897	0.448	0.370	-0.793
			OTD	19.791	18.495	17.719	0.484	0.444	-0.836
	SCA-DSF		LTH	19.166	15.630	17.333	0.528	0.433	-0.812
			OFC	15.530	12.176	12.700	0.442	0.365	-0.776
			OTD	19.075	17.267	17.290	0.488	0.451	-0.819
	Input Mixture		LTH	-9.741	0.045	-6.416	0.251	0.173	0.979
			OFC	-8.231	0.045	-2.895	0.311	0.229	1.000
			OTD	-9.370	0.205	-6.438	0.297	0.171	0.982

Table 3-1: Performance summary for two-source separation
LTH - lecture theatre; OFC - office; OTD - outdoor.

The permissions to perform any process with such functionalities must be declared in the manifest as shown in Figure 3-1.

```xml
<?xml version="1.0" encoding="utf-8"?>
<manifest xmlns:android="http://schemas.android.com/apk/res/android"
    <!-- Package declaration...  -->

    <uses-permission android:name="android.permission.RECORD_AUDIO" />
    <uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE" />
    <uses-permission android:name="android.permission.INTERNET" />
    <uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" />

    <!-- The rest of the manifest... -->
</manifest>
```

Figure 3-1: Declaration of the permissions required in the Android manifest

3.1.1 Recording of the Observed Mixture

As the user presses the record button (A), the method `startRecording()` is invoked and the record button's icon changes to a stop symbol as shown in Figure 3-2.

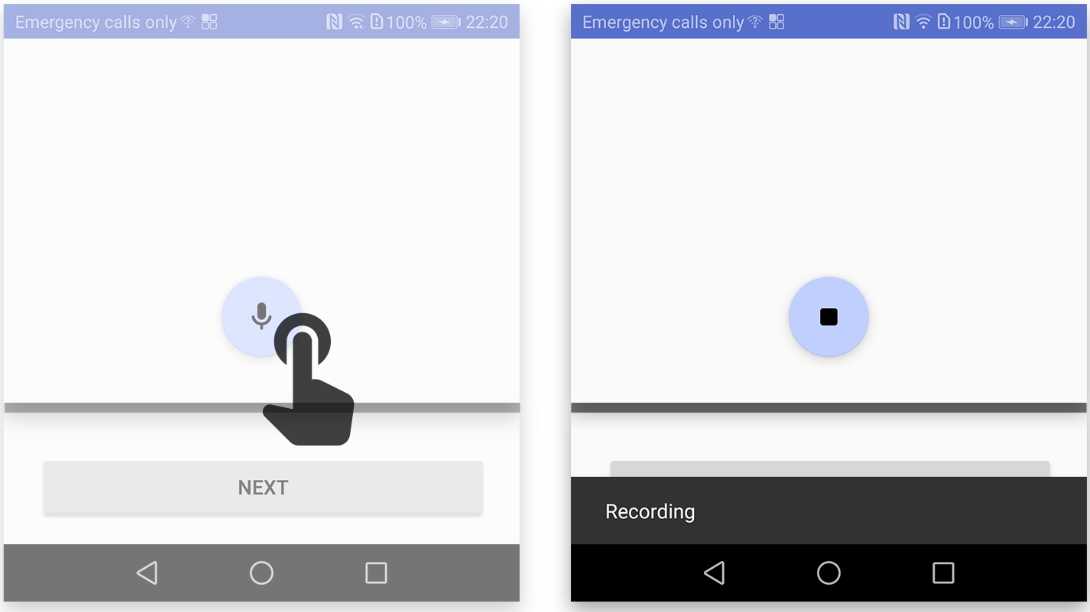


Figure 3-2: Record button icon changes from a microphone to a stop symbol

4. Conclusion and Future Work

4.1 Project Summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

4.2 Areas of Improvement

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

4.3 Future Works

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

References

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83.

Sapp, C. S. (2004). WAVE PCM soundfile format. http://soundfile.sapp.org/doc/WaveFormat/

Scheibler, R. (2018). Common functions used in BSS algorithms [MIT License]. https://github.com/LCAV/pyroomacoustics/blob/pypi-release/pyroomacoustics/bss/common.py

Smaragdis, P. (1998). Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, *22*(1-3), 21–34.

Smith III, J. O. (2011). *Spectral audio signal processing*. W3K publishing.

Appendix Tables and Figures

Table A-1

Table A-1 shows the speaker initials (ID), sex, dialect region number (DR), and the sentences used for each of the sources. Sentences that are trimmed are denoted with asterisks.

Source	Speaker ID	Sex	DR	Sentences			
1	JWT0	M	1	SI1291	SI751	SI1381*	
2	AEM0	F	2	SA1	SA2	SI762	SI1392*
3	SLS0	F	3	SI1056	SI1686	SI2316	
4	BAS0	F	4	SI1387	SI1472	SI2066*	
5	DWH0	M	5	SI1168	SI1925	SX35	
6	JRK0	M	6	SI1662	SI2130	SI880	SX160*

Table A-1: Speaker information for the evaluation sources
M - male; F - female. DR1 - New England; DR2 - Northern; DR3 - North Midland; DR4 - South Midland; DR5 - Southern; DR6 - New York City.

Table A-2

Source	Transcript
1	they should live in modest circumstances avoiding all conspicuous consumption serve in frankfurter buns or as a meat dish but briefly the topping
2	she had your dark suit in greasy wash water all year don't ask me to carry an oily rag like that fill small hole in bowl with clay assume for
3	can thermonuclear war be set off by accident it latches when you close it so stay as long as you like Davy Mathews it's disgusting the way you're always eating
4	several factors contributed to this change she greeted her husband's colleagues with smiling politeness offering nothing He saw a pint-sized man
5	it takes a great deal of sophisticated thought to get the impact of this fact so what's this all about help celebrate your brother's success
6	did anyone see my cab See you in about an hour the revolution now under way in materials handling makes this much easier as co-authors we presented our new book

Table A-2: Transcript of the evaluation sources

Figure A-1

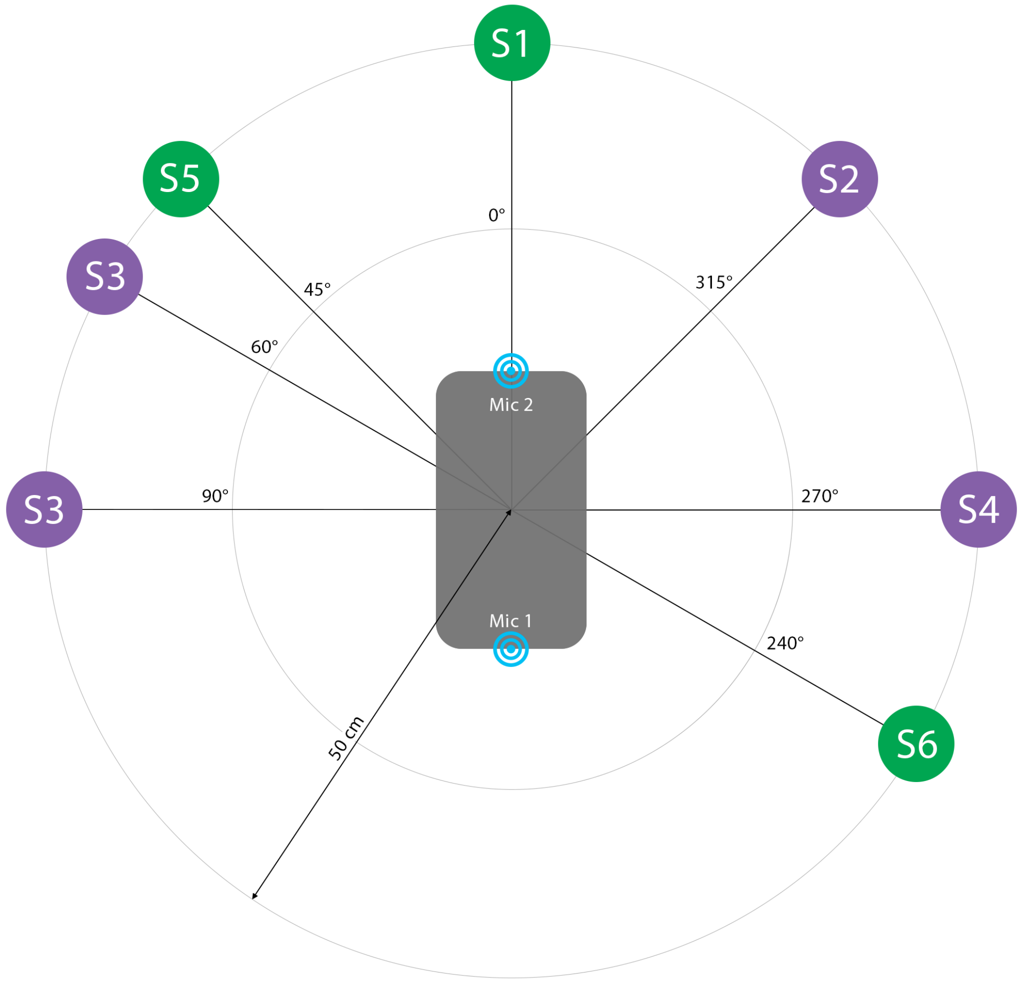


Figure A-1: Positions of the loudspeaker for each source speaker
A green circle denotes a male speaker. A purple circle denotes a female speaker. The angle is measured with respect to the line passing through the microphones in a counterclockwise manner, starting from the top of the phone where Microphone 2 is located. Note that Speaker 3 was recorded at two different positions. The figure is not to scale.

Figures A-2 - A5

The configurations of the mixtures are shown, not to scale, in Figures A-2 to A-5. The source configuration for each mixture is the same across all acoustic environments.

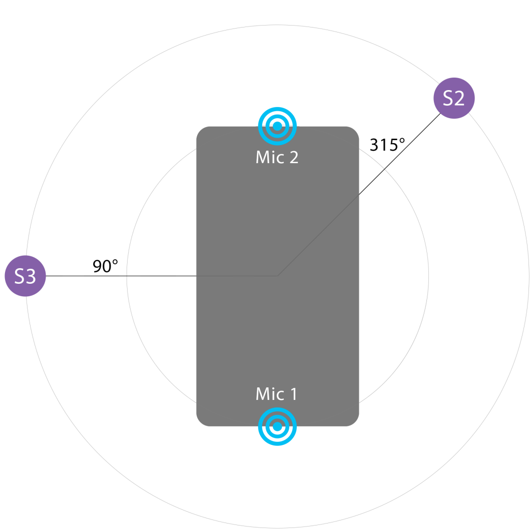


Figure A-2: Configuration 1 for two-source mixture

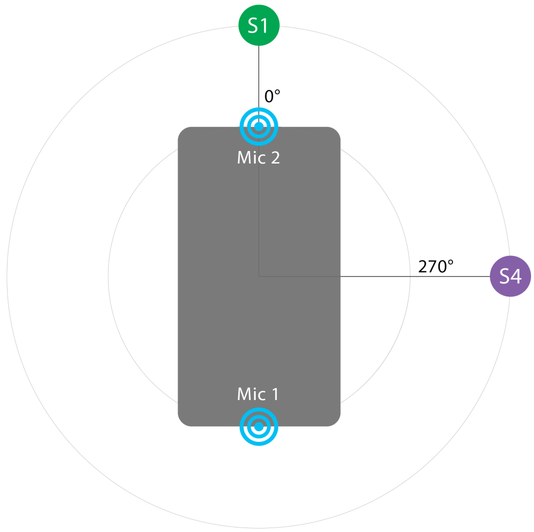


Figure A-4: Configuration 1 for three-source mixture

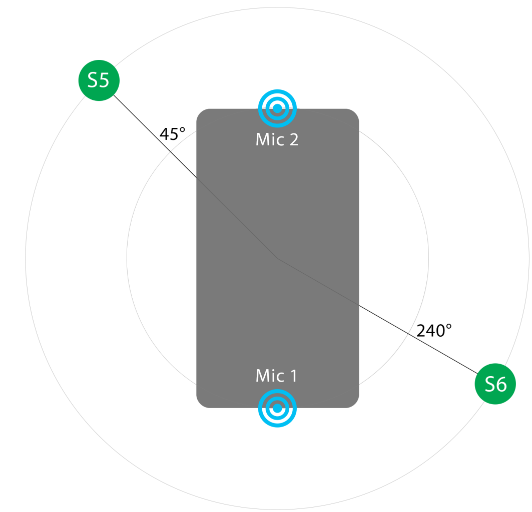


Figure A-3: Configuration 2 for two-source mixture

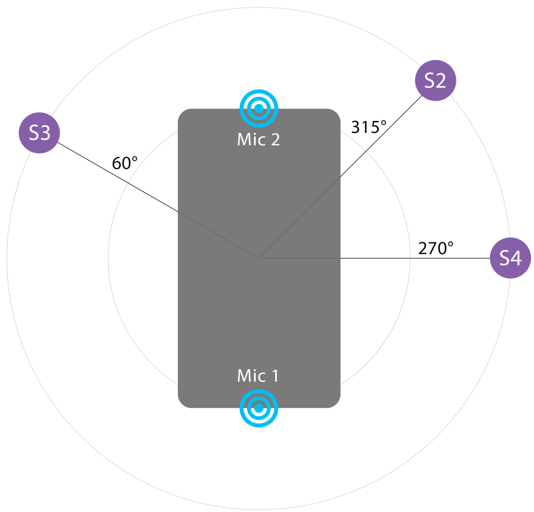


Figure A-5: Configuration 2 for three-source mixture